u^b

b

**UNIVERSITÄT
BERN**

Faculty of Economics and
Social Sciences

Department of Economics

**On the causal effect of
schooling on smoking:
evidence without exogeneity conditions**

Stefan Boes

11-02

April 2011

DISCUSSION PAPERS

On the causal effect of schooling on smoking: evidence without exogeneity conditions[*]

Stefan Boes[†]

April 7, 2011

Abstract

The paper explores weak monotonicity and convexity assumptions in a model for the decision to smoke with endogenous schooling. Theories of productive and allocative efficiency as well as the influence of time preferences are accounted for in order to derive testable constraints that bound the effect of schooling on smoking. Data from the Swiss Health Survey indicate that the degree of endogeneity depends on the level of schooling, and that schooling effects are likely heterogeneous with a reduction of the propensity to smoke by at most 5.9 percentage points.

JEL Classification: I12, C14, C30

Keywords: Smoking, education, health behavior, nonparametric bounds.

[†]Correspondence: University of Bern, Department of Economics, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland, phone: +41 31 631 4792, fax: +41 31 631 3783, email: stefan.boes@vwi.unibe.ch.

1 Introduction

The effect of schooling on non-market outcomes has received interest of the economics profession at least since Becker's work on the allocation of time and production of non-market goods (Becker, 1960, 1965, 1991). Schooling typically enters the production process through efficiency parameters and thereby affects the marginal costs of the non-traded good (Rosenzweig and Schultz, 1989; Grossman, 1972). Theories consistent with this view justify a causal schooling effect. At the same time, the literature has recognized the potential endogeneity of schooling. Preferences and tastes, for example, may play the role of confounding factors in the production of goods and in the determination of schooling (Fuchs, 1982; Farrell and Fuchs, 1982). Grossman (2006) surveys the related theoretical and empirical literature.

The aim of this paper is to explore both types of theories in order to estimate the causal effect of schooling on smoking. Previous evidence based on exogenous variation has documented a negative and significant schooling effect (Sander 1995a,b; Currie and Moretti, 2003; Kenkel et al., 2006; de Walque, 2007; Grimard and Parent, 2007; Gilman et al., 2008; Tenn et al., 2008). However, credible information on instruments may not always be available. The paper demonstrates how substantive estimates of the causal effect can be obtained even in the absence of such variation. On the one hand, theories of productive and allocative efficiency, following the categorization of Grossman (2006), impose functional restrictions on the propensity to smoke with increased schooling. On the other hand, the third-variable hypothesis restricts the way how the propensity to smoke and education decisions jointly vary in subgroups of the population.

Two conclusions can be drawn from the theoretical models. First, they unambigu-

ously predict the signs of effects, and thus can be used to impose monotonicity constraints on the observed data distribution. In addition, the paper makes explicit use of supplementary information that is common to many economic models, namely that of decreasing marginal returns (e.g., Kreps, 1990: Ch. 2; Levhari and Weiss, 1974; Moffitt, 2007). For example, it may be plausible that additional knowledge gains are smaller the higher the level of education, and that the marginal gains of health care usage increase at a diminishing rate. Given a negative impact of education, this translates into convexity constraints that add information to the identification strategy.

The contributions of the paper are twofold. First, it formally outlines how the convexity aspect can be accounted for in testable restrictions on the data-generating process. As a main result, the causal effect of schooling on the propensity to smoke can only be bounded, with the width of bounds depending on the strength of the underlying assumptions (Manski and Pepper, 2000; Boes, 2010). Second, the paper shows that monotonicity and/or convexity may provide substantive estimates of the causal schooling effect, even without imposing exogeneity conditions. Using data from the Swiss Health Survey, the analysis suggests that each additional year of schooling reduces the propensity to smoke by at most 5.9 percentage points.

The paper proceeds as follows. The theoretical background and the assumptions on the data generating process are provided next. The potential outcomes framework will be used to discuss identification and estimation. The particularities of the data are described in the third section. Sections 4 and 5 report the results, first the bounds on the potential outcome distribution, and second the bounds on the causal effect of schooling on the propensity to smoke. Section 6 concludes.

2 Theoretical background

Several theories explain the productive efficiency effect of schooling on health (or non-market outcomes in general), e.g., Michael (1972, 1973), Michael and Becker (1973), and Grossman (1972, 1975, 2000, 2006). Productive efficiency stresses that people with higher education translate health inputs (such as medical care or time) into outputs (such as smoking) in a more efficient way because education raises the marginal products of the inputs. I will use the static version of the Grossman (1972) pure investment model, as discussed in Grossman (2006), to illustrate this effect. Suppose that in a given period, say one year, the amount of time (h) allocated to market and non-market production is an increasing function of health (H), for example because better health lowers the risk of illness and the time lost through it. Moreover, assume that the marginal product of health falls due to a finite upper limit of time available for the production of market and non-market goods, so $\partial h / \partial H > 0$, $\partial^2 h / \partial H^2 < 0$.

The production of health is characterized by the following process

$$H = e^{\rho(S)} F(M, T) \tag{1}$$

where F is linear homogeneous in the inputs medical care (M) and own time of the consumer (T). Schooling (S) is assumed to affect the marginal products of M and T by the same percentage $\rho(S)$, where $\rho(S)$ is an increasing function of S. This is a factor neutrality assumption, and hence increases in S do not change the relative equilibrium use of M and T. The marginal products of M and T are assumed to have a finite upper limit in S because at some stage everything is known about their usage in the production of health, and thus $\partial \rho(S) / \partial S > 0$, $\partial^2 \rho(S) / \partial S^2 < 0$. Consumers maximize net benefits $Wh - \pi_H H$, where W is the wage rate and π_H is the marginal or average

cost of producing health. Optimal demand for health is determined by the equality of marginal benefits and marginal costs, i.e., $W \frac{\partial h}{\partial H} = \pi_H$. As in Grossman (2000, 2006), a unit increase in schooling S causes first order changes in optimal H by

$$\frac{dH}{dS} = -\frac{\partial \rho(S)}{\partial S} \frac{\partial h/\partial H}{\partial^2 h/\partial H^2} > 0 \tag{2}$$

Given that individual time preferences are fixed, the second order changes in the demand for health are given by

$$\frac{d^2 H}{dS^2} = -\frac{\partial^2 \rho(S)}{\partial S^2} \frac{\partial h/\partial H}{\partial^2 h/\partial H^2} < 0 \tag{3}$$

and thus, increases in health are diminishing with additional schooling.

Theories of allocative efficiency stress that the better educated individuals choose a more efficient mix of inputs, because they have, for instance, additional knowledge about the use of inputs, which ultimately gives them a larger amount of the non-market good. Unlike for the productive efficiency channel, education affects outcomes only if the mix of inputs is changed (e.g., Rosenzweig and Schultz, 1983, 1989; Kenkel, 1991; Rosenzweig, 1995; Goldman and Lakdawalla, 2005).

In the following, I will illustrate the allocative efficiency argument using a model outlined in Glied and Lleras-Muney (2008). Suppose that health is produced using technology A and other inputs C. The level of technology is assumed to grow at a constant rate λ, and the technology available to each individual depends on the time between innovation and adoption. The latter is assumed to depend on the level of education in a way that

$$A(t) = T_0 e^{\lambda(t-w(S))} \tag{4}$$

where T_0 is the frontier level of technology at time $t = 0$. $w(S)$ is assumed to decrease in schooling S because the better educated adopt new technologies at a faster rate than the less educated (e.g., due to a better access to and use of information in the production of health). Moreover, this rate is assumed to decrease in schooling because the time of adoption cannot occur prior to the innovation, so that $\partial w(S)/\partial S < 0$ and $\partial^2 w(S)/\partial S^2 > 0$. Given a health production function $H = H(A, C)$ with $\partial H(A, C)/\partial A > 0$, the impact of education on health can be evaluated as

$$\frac{dH}{dS} = -\lambda A \frac{\partial w(S)}{\partial S} \frac{\partial H(A, C)}{\partial A} > 0 \tag{5}$$

$$\frac{d^2 H}{dS^2} = -\lambda A \frac{\partial^2 w(S)}{\partial S^2} \frac{\partial H(A, C)}{\partial A} < 0 \tag{6}$$

which yields a decreasing marginal effect of education on health.

The third-variable hypothesis highlights a different aspect of the education-health gradient. Fuchs (1982) and Farrell and Fuchs (1982), for example, argue that time preferences affect schooling decisions as well as health related behavior. People who are more future oriented, i.e., who discount time at a lower rate, will invest more in schooling because they expect higher returns in the future. But people with a lower time discount will also make larger investments in their own health because they expect to live longer and/or they draw positive utility from future activities. De Walque (2007b) discusses this view in a specific model for health related behavior with different types of individuals with varying stock of human capital (see also Grossman, 2006)

At the core of the time preferences argument is the maximization of an intertemporal utility function defined as the discounted sum of utilities U_t at each age t

$$U = \sum_{t=0}^{T} D^t U_t \tag{7}$$

where D is the discount factor. If consumers are more present (future) oriented, then D will be small (large). De Walque (2007b) and Grossman (2006) suggest that the relationship between schooling and health can be characterized by

$$\frac{dH}{dS} = \frac{dH/dD}{dS/dD} > 0 \tag{8}$$

because $dH/dD > 0$ and $dS/dD > 0$.

A sufficient condition for the second order changes to be negative would be that increases in education are larger than increases in health if the time discount increases. The evaluation of the relative impact of the time discount is ultimately an empirical question, and a number of recent papers deal with this subject (Cairns, 2006; Khwaja et al. 2006; Scharff and Viscusi, 2010). In a context similar to this paper, van der Pol (2010) finds an almost linear impact of the time discount on health, and a convex impact on schooling (see Table II in van der Pol, 2010). This suggests that $d^2H/dS^2 < 0$.

Now suppose that smoking status is one possible indicator of health. Let $Y(s)$ denote the potential smoking status ($0 = $ non-smoker, $1 = $ smoker) for each conjectured schooling level s, i.e., if the individual was exogenously assigned schooling $s \in \mathcal{S}$. Theories of productive and allocative efficiency imply that the propensity to smoke decreases at a diminishing rate with additional schooling s. This holds for each subgroup of the population, with any realized schooling $S = t$. Formally,

Assumption (1). *For all $(s_1, s_2, t) \in \mathcal{S}^3$ with $s_1 \leq s_2$, and for all $\alpha \in [0,1]$, (i)*
$$P[Y(\alpha s_1 + (1-\alpha)s_2) = 1|S = t] \leq \alpha P[Y(s_1) = 1|S = t] + (1-\alpha)P[Y(s_2) = 1|S = t],$$
and (ii) $P[Y(s_1) = 1|S = t] \geq P[Y(s_2) = 1|S = t]$.

Part (i) accounts for the convexity aspect, part (ii) implies a weakly monotonically decreasing propensity to smoke if schooling is exogenously increased. The implications

7

of assumption (1) for the identification of the propensity to smoke if everybody was assigned s years of schooling are discussed in Boes (2010); see also Manski (1997) and Manski and Pepper (2000). In a nutshell, assumption (1) allows to combine (by (i)) the observed propensity to smoke for those with $S = t$, $P[Y = 1|S = t]$, with the upper bounds of the counterfactual probabilities $P[Y(s) = 1|S = t]$ for all $s \neq t$ that follow from the monotonicity (ii) of $P[Y(s) = 1|S = t]$ in s.

The third-variable hypothesis translates into a somewhat different assumption. As outlined above, people with a larger time discount D are characterized by a higher realized schooling and a lower propensity to smoke (due to their different health behavior). Under convexity the assumption can be stated formally as

Assumption (2). *For all $(s, t_1, t_2) \in \mathcal{S}^3$ with $t_1 \leq t_2$, and for all $\beta \in [0, 1]$, (i)*

$$P[Y(s) = 1|S = \beta t_1 + (1 - \beta)t_2] \leq \beta P[Y(s) = 1|S = t_1] + (1 - \beta)P[Y(s) = 1|S = t_2],$$

and (ii) $P[Y(s) = 1|S = t_1] \geq P[Y(s) = 1|S = t_2]$.

The identifying power of assumption (2), and assumptions (1) and (2) jointly is outlined in Boes (2010). Like before, assumption (2) allows to make linear combinations of the observed propensity to smoke $P[Y = 1|S = s]$ with the upper bounds on $P[Y(s) = 1|S = t]$ for all $t \neq s$. If both assumptions are imposed jointly, then this allows to linearly combine all observed propensities to smoke for each schooling level to find a sharp upper bound on the effect of schooling on smoking (basically by estimating the lowest convex combination of observed data points, see the calculations below). Moreover, the convexity in both directions (s and t) implies convexity of the observed data distribution $P[Y = 1|S]$ in realized schooling S, which is a testable assumption.

3   Data

The analysis is based on the 2007 wave of the Swiss Health Survey (SHS). The SHS is a representative cross section of the Swiss population that records a large number of socioeconomic and health related characteristics. For this study, I extracted data on tobacco consumption (yes/no; which includes smoking of cigarettes, cigars, and/or pipes), information on the highest education achieved as well as the respondent's age, gender, and monthly net personal income. Raw education is stored in 10 categories which are linked to the average number of formal years of schooling using official school statistics (see the appendix). The final sample consists of 10,189 individuals after consistency checks, cleaning for missing values and invalid responses.

Table 1 reports the relative proportion of smokers by the years of schooling. The simple bivariate statistics suggest that the propensity to smoke is highest among the less educated, and smallest among the more educated individuals. Differences in the propensity smoke by schooling are larger the smaller the level of schooling which points to a non-linear, possibly convex association.

No causal conclusion can be drawn from Table 1, though, as the relationship between schooling and smoking is likely confounded by observed and unobserved characteristics. Possible background factors are the respondent's age and gender. Figure 1 regression adjusts for the latter two and plots the relative number of smokers for each schooling category, given an average age of 48.8 years and a sample proportion of men of 46 percent. The adjusted results still suggest a downward sloping relationship between the years of schooling and the propensity to smoke. Using polynomial smoothing, the best fit is achieved if education enters the regression function in a quadratic form (indicated

by the dashed line). While the observed data points do not display a perfectly convex shape, the hypothesis of a convex relationship cannot be rejected.

In order to assess the causal effect of schooling on smoking, one might simply assume that schooling is determined exogenously, perhaps conditional on age and gender. In this case, Figure 1 would be sufficient to obtain the desired effect. For example, an increase in schooling from 13 to 15 years was predicted to decrease the propensity to smoke from 23.9 to about 21.3 percent, a difference of about 2.6 percentage points. However, the exogeneity of schooling is not a plausible assumption in the presence of unobserved confounders (such as time preferences or tastes) that dilute the causal interpretation of Figure 1. The next two sections therefore exploit the assumptions stated above in order to quantify the causal schooling effect.

4   Bounds on the distribution of potential outcomes

Using the empirical evidence alone, the potential outcome distribution $P[Y(s)]$ can be bounded by $P(Y|S = s)P(S = s) + \gamma P(S \neq s)$, with $\gamma = 0$ for the lower bound, and $\gamma = 1$ for the upper bound. In many cases, as they are in this application, these no assumptions bounds (Manski, 1990) are fairly wide and do not allow for informative conclusions. In particular, due to the many schooling categories, the probability of observing s is low relative to not observing s (with the exception of one category; see Table 1), and hence, the obtained lower bounds are close to zero and the obtained upper bounds are close to one, which is about the logical range for a probability function.

Assumptions (1) and (2) impose restrictions on γ such that tighter bounds on the potential outcome distribution can be obtained. I will explore these assumptions in three

steps, with increasing strength of identifying power. First, the monotonicity conditions as stated in (1, ii) and (2, ii) will be imposed. Second, convexity in the sense of (1, i) will be explored as additional restriction. Third, convexity as implied by (1) and (2) jointly will be used to constrain the observed data distribution.

Assumptions (1, ii) and (2, ii) together imply a monotonically decreasing propensity to smoke, as a function of schooling. The observed data points (see Figure 1) are consistent with this implication, and thus the evidence cannot reject the joint monotonicity hypothesis. Moreover, $P(Y = 1|S = t)$ can be used as a tight lower bound and $P(Y = 1|S = s)$ as a tight upper bound on $P[Y(s) = 1|S = t]$ for all $s > t$, and vice versa for all $s < t$. These bounds can be used to derive the bounds on the potential outcome distribution for each schooling level as indicated in Figure 2. Monotonicity provides substantive information on the potential propensity to smoke by restricting the range of possible values to a width of at most 5 percentage points (except for the smallest education level).

If the propensity to smoke decreases at a diminishing rate with additional schooling (assumption (1)), and (2, ii) is still fulfilled, then the tightest convex combination of observed data points can be used as a tight upper bound on $P[Y(s) = 1|S = t]$ for all $s < t$. For example, the linear combination $0.5 \cdot P(Y = 1|S = 9) + 0.5 \cdot P(Y = 1|S = 13)$ is smaller than $P(Y = 1|S = 11)$ and thus can be used to further improve upon the monotonicity upper bound on $P[Y(11) = 1|S = 12]$. Such improvements are possible for education levels 11, 12, 16, and 18. The resulting bounds are shown as the middle bars in Figure 3 (left bars indicate the joint monotonicity of Figure 2). As it can be seen, there are modest gains on the width compared to the monotonicity bounds.

11

Convexity in the sense of (1) and (2) allows for an extension of the upper bounds argument through combinations of the observed data points to the case $s > t$. The bounds under this scenario are shown by the right bars in Figure 3. The width shrinks to at most 3.5 percentage points, except for the lowest/highest education levels which are not affected. The range of potential outcome probabilities varies by education. For example, if everybody was assigned 12 (15) years of schooling, then the propensity to smoke would be predicted 28.5 to 30.0 percent (24.8 to 28.7 percent). Figure 3 also shows that the observed data points are in most cases inside the stated bounds, but are outside for 11 and 12 years of schooling. The exogeneity of schooling is therefore inconsistent with the convexity constraints for these two levels.

If personal income is held constant, then the results do not change much. Figure 4 shows that the obtained bounds under the three sets of assumptions are almost the same as in Figure 3. Two effects are at play. First, the regression adjusted point estimates of the education dummies are reduced by about 8-10 percentage points if income is fixed, consistent with the previous empirical evidence cited above. Second, if adjusted for the mean value of the logarithmic income variable of about 7.6 and estimated coefficient 0.013, the reduction is almost entirely compensated. Since the functional form relationship between schooling and smoking does not change conditional on income, the range of possible values for the potential propensity to smoke is not sensitive to the income variable being fixed. In terms of a parsimonious model specification, the income variable is therefore not considered further. This result is unlike the respondents' age and gender that both change the functional relationship if not conditioned on.

5 Bounds on the causal effect

The implications of assumptions (1) and (2) regarding the causal schooling effect, formally defined as $P[Y(s_2) = 1] - P[Y(s_1) = 1]$ for $s_2 > s_1$, are not directly inferable from Figures 2 and 3. A simple comparison of upper bounds on $P[Y(s_1) = 1]$ and lower bounds on $P[Y(s_2) = 1]$ in order to obtain a lower bound on the causal effect is invalid because the overall bounds are obtained as a weighted average of upper/lower bounds, and thus the weights (in this case the schooling distribution) matter. Table 2 provides valid estimates for the interventions indicated in column 1.

Under exogeneity, the causal effect is point identified and equal to the observed difference between the propensities to smoke for adjacent schooling levels (column 2). These differences are inferable from Figure 1 (by comparison of the dots). Column 3 shows the Manski (1990) no assumptions bounds, i.e., the bounds that are obtained under no restrictions on the data generating process. As can be seen, the latter are very wide and allow for almost any admissible negative effect. This result was expected given the discussion above.

Like in the preceding section, the monotonicity/convexity assumptions are explored in three steps: (i) monotonicity only, (ii) convexity from (1) and monotonicity from (2), and (iii) convexity from (1) and (2). Imposing monotonicity substantially increases the lower bound on the causal effect and thus decreases the width of the bounds, given the fixed upper bound zero (column 4). The point estimates vary between -11.8 percentage points and -3.4 percentage points. Since the estimates are based on different schooling effects, one might divide by the absolute increase in schooling in order to get a comparable lower bound for one-year raises. The results suggest that this lower bound lies

13

between -1.5 and -4.6 percentage points (at most -5.9 percentage points for an additional year of schooling between 7 and 9 years).

If convexity in the sense of (1, i) is imposed as additional assumption (column 5), then the width of the bounds further shrinks, and if convexity in the sense of (1) and (2) is imposed jointly (column 6), then additional gains in terms of more informative point estimates are obtained. Under the latter set of assumptions, the estimated lower bound lies between -1.3 and -3.9 percentage points (again with the exception of the lowest years of schooling which are not affected by the convexity constraints). These results are consistent with the previous evidence using instrumental variables, e.g., Currie and Moretti (2003), de Walque (2007a), Grimard and Parent (2007), and Tenn et al. (2010). It should be noted that the observed differences are outside the estimated bounds for two scenarios (15 → 16 and 16 → 17), and thus again, the results suggest that exogeneity cannot be rejected over the entire schooling range.

Additional improvements can be obtained if the results using instrumental variables are taken into account. In specifications similar to the one chosen here, Currie and Moretti (2003), de Walque (2007a) and Grimard and Parent (2007) show that the instrumental variables estimates are larger in absolute magnitude than the ordinary least squares estimates. If measurement error is small in magnitude and the effects are constant within educational levels, then the OLS estimates may be used as an upper bound on the percentage point change instead of zero. These upper bounds are given in Table 2, column 2. The results suggest that the effect of increasing schooling from 12 to 13 years, for example, reduces the propensity to smoke by about 2.3 to 3.4 percentage points, a reasonably small range of values to draw informative policy conclusions.

Table 2 can also be used to obtain the lower bound on a different causal effect, namely the effect of increasing the years of schooling to the next higher level for subgroups with a given education. Formally, this effect can be expressed as $P[Y(s_2) = 1|S = s_1] - P[Y(s_1) = 1|S = s_1]$, which is often referred to as the treatment effect on the treated. The latter term is identified from the observed data distribution by the propensity to smoke for those with s_1 years of schooling. A lower bound on the former term can be obtained as $P[Y = 1|S = s_2]$ under the monotonicity constraints. Thus, the lower bounds on the treatment effect on the treated are given by column 2, with the positive estimates set to zero. The results suggest that this effect is at most -2.3 percentage points per additional year of schooling, again with the exception of the lowest schooling levels with -4.1 percentage points for the estimated lower bound.

6   Conclusion

The adverse health effects of smoking are well-known and reach from increased risk of lung cancer and coronary heart disease to adverse reproductive and early childhood effects (among others, US Department of Health and Human Services, 2004). If people invest in their education at sub-optimal levels (evidence in favor of this is provided, for example, by Levhari and Weiss (1974)), and schooling has a causal effect on smoking, then there are good reasons to subsidize individual skill formation, which will ultimately improve health and related behaviors.

This paper used weak monotonicity and convexity assumptions in the spirit of Manski and Pepper (2000) and Boes (2010) to bound the causal effect of schooling on smoking. The assumptions were motivated in the context of economic theories of productive and

15

allocative efficiency accounting for the endogeneity of schooling through unobserved time preferences. The results suggest that the magnitude of effects depends on the schooling level. The lowest bound of -5.9 percentage points is obtained for the lowest educated individuals, the estimates for the other education groups suggest a bound of at most -3.9 percentage points. The results are consistent with the previous literature using instrumental variables, but the analyses here do not require exogenous variation in order to obtain substantive estimates of causal education effects.

The analysis could be extended to allow for variation in subgroups. Cutler and Lleras-Muney (2008, 2010), for example, use different data sources to quantify the relative importance of various explanations for the education gradient. Using the proposed identification strategy one would condition on covariates or regression-adjust the observed propensity to smoke as a function of schooling. In that way, the relative impact can be evaluated. However, the SHS is not large enough to allow for such detail because the propensity to smoke needs to be estimated for each education level. It would therefore be reasonable to allow for some smoothing over the schooling range.

References

Becker GS, 1960. An economic analysis of fertility. In: Demographic and economic change in developed countries, Princeton University Press, p. 209-231.

Becker GS, 1965. A theory of the allocation of time. Economic Journal 75, 493-517.

Becker GS, 1991. A treatise on the family, 2nd edition. Harvard University Press.

Boes S, 2010. Convex treatment response and treatment selection. SOI Working Paper 1001, University of Zurich.

Cairns J, 2006. Developments in discounting: with special reference to future health events. Resource and Energy Economics 28, 282-297.

Currie J, Moretti E, 2003. Mothers education and the intergenerational transmission of human capital: evidence from college openings. Quarterly Journal of Economics 118, 1495-1532.

Cutler DM, Lleras-Muney A, 2008. Education and health: evaluating theories and evidence. In: Schoeni RF, House JS, Kaplan G, Pollack H (eds) Making americans healthier: social and economics policy as health policy, Russell Sage Foundation.

Cutler DM, Lleras-Muney A, 2010. Understanding differences in health behaviors by education. Journal of Health Economics 29, 1-28.

de Walque D, 2007a. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. Journal of Health Economics 26, 877-895.

de Walque D, 2007b. How does the impact of an HIV/AIDS information campaign vary with educational attainment? Evidence from rural Uganda. Journal of Development Economics 84, 686-714.

Fuchs VR, 1982. Time preference and health: an exploratory study. In: Fuchs VR (ed) Economic aspects of health, University of Chicago Press, p. 93-120.

Farrell P, Fuchs VR, 1982. Schooling and health: the cigarette connection. Journal of Health Economics 1, 217-230.

Glied S, Lleras-Muney A, 2008. Health inequality, education and medical innovation. Demography 45, 741-761.

Gilman SE, Martin LT, Abrams DB, Kawachi I, Kubzansky L, Loucks EB, Rende R, Rudd R, Buka SL, 2008. Educational attainment and cigarette smoking: a causal association? International Journal of Epidemiology 37, 615-624.

Goldman DP, Lakdawalla DN, 2005. A theory of health disparities and medical technology. Contributions to Economic Analysis and Policy 4, Article 8.

Grimard F, Parent D, 2007. Education and smoking: were Vietnam draft avoiders also more likely to avoid smoking? Journal of Health Economics 26, 896-926.

Grossman M, 1972. On the concept of health capital and the demand for health. Journal of Political Economy 80, 223-255.

Grossman M, 1975. The correlation between health and schooling. In: Terleckyj NE (ed) Household production and consumption, Columbia University Press, p. 147-211.

Grossman M, 2000. The human capital model. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics 1A, Elsevier, p. 347-408.

Grossman M, 2006. Education and non-market outcomes. In: Hanushek E, Welch F (eds) Handbook of the economics of education 1, Elsevier, p. 577-633.

Kenkel DS, 1991. Health behavior, health knowledge, and schooling. Journal of Political Economy 99, 287-305.

Kenkel DS, Lillard D, Mathios A, 2006. The roles of high school completion and GED receipt in smoking and obesity. Journal of Labor Economics 24, 635-660.

Khwaja A, Sloan F, Salm M, 2006. Evidence on preferences and subjective beliefs of risk takers: the case of smokers. International Journal of Industrial Organization 24, 667-682.

Kreps DM, 1990. A course in microeconomic theory, Princeton University Press.

Levhari D, Weiss Y, 1974. The effect of risk on the investment in human capital. American Economic Review 64, 950-963.

Manski CF, 1990. Nonparametric bounds on treatment effects. American Economic Review, Papers and Proceedings 80, 319-323.

Manski CF, 1997. Monotone treatment response. Econometrica 65, 1311-1334.

Manski CF, Pepper J, 2000. Monotone instrumental variables: with an application to the returns to schooling. Econometrica 68, 997-1010.

Michael RT, 1972. The effect of education on efficiency in consumption. Columbia University Press.

Michael RT, 1973. Education in non-market production. Journal of Political Economy 81, 306-327.

Michael RT, Becker GS, 1973. On the new theory of consumer behavior. The Swedish Journal of Economics 75, 378-396.

Moffitt RA, 2007. Estimating the marginal returns to higher education in the UK, NBER Working Paper 13534.

Rosenzweig MR, 1995. Why are there returns to schooling? American Economic Review 85, 153-158.

Rosenzweig MR, Schultz TP, 1983. Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight. Journal of Political Economy 91, 723-746.

Rosenzweig MR, Schultz TP, 1989. Schooling, information and nonmarket productivity: contraceptive use and its effectiveness. International Economic Review 30, 457-477.

Sander W, 1995a. Schooling and quitting smoking. Review of Economics and Statistics 77, 191-199.

Sander W, 1995b. Schooling and smoking. Economics of Education Review 14, 23-33.

Scharff RL, Viscusi WK, 2010. Heterogenous rates of time preference and the decision to smoke. Economic Inquiry, forthcoming.

Tenn S, Herman DA, Wendling B, 2010. The role of education in the production of health: an empirical analysis of smoking behavior. Journal of Health Economics 29, 404-417.

US Department of Health and Human Services, 2004. The health consequences of smoking: a report of the surgeon general. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health
[last accessed 2010, Oct 27].

Van der Pol M, 2010. Health, education and time preference. Health Economics, forthcoming.

# Tables and Figures

Table 1: Summary statistics conditional on schooling

| Schooling (S)	$P(S)$	$P(smoke|S)$	$P(male|S)$	$E(age|S)$	$E(inc|S)$	N
7 years	0.0044	0.400	0.42	51.0	7.02	45
9	0.0947	0.301	0.32	53.3	6.87	966
11	0.5456	0.309	0.43	49.2	7.39	5,564
12	0.0293	0.294	0.38	48.0	7.32	299
13	0.0095	0.258	0.10	45.6	6.95	97
15	0.0947	0.257	0.53	48.2	7.98	966
16	0.0950	0.271	0.63	46.8	8.37	969
18	0.0706	0.275	0.56	45.9	8.41	720
22	0.0561	0.251	0.62	47.0	8.53	572
Total	1	0.293	0.46	48.8	7.62	10,198

Source: Swiss Health Survey 2007, own calculations. *Notes on variables: smoke* equals one if the respondent consumes cigarettes, cigars, and/or pipes. *inc* is the log of monthly personal net income in Swiss Francs.

Figure 1: Adjusted propensity to smoke conditional on schooling
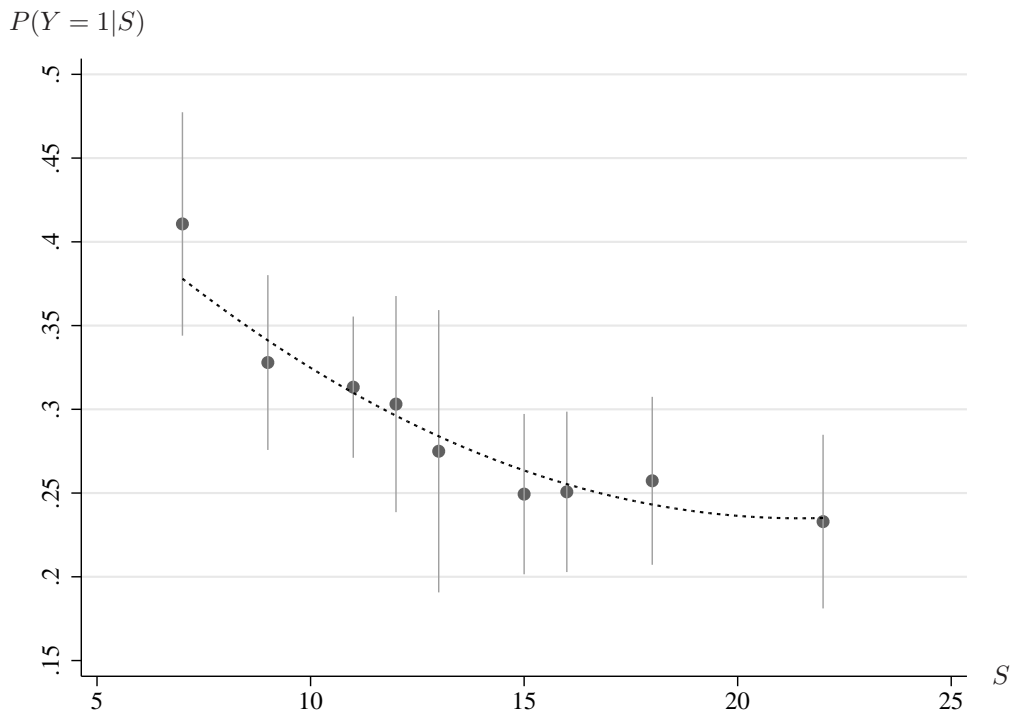
Source: Swiss Health Survey 2007, own calculations. *Notes:* The propensity to smoke (Y) for each schooling level (S) is regression adjusted for the respondents' age and gender. Point estimates are shown by the black dots, the vertical grey bars display a 95% confidence interval. The dashed line is a smoothed quadratic fit over the observed data points.
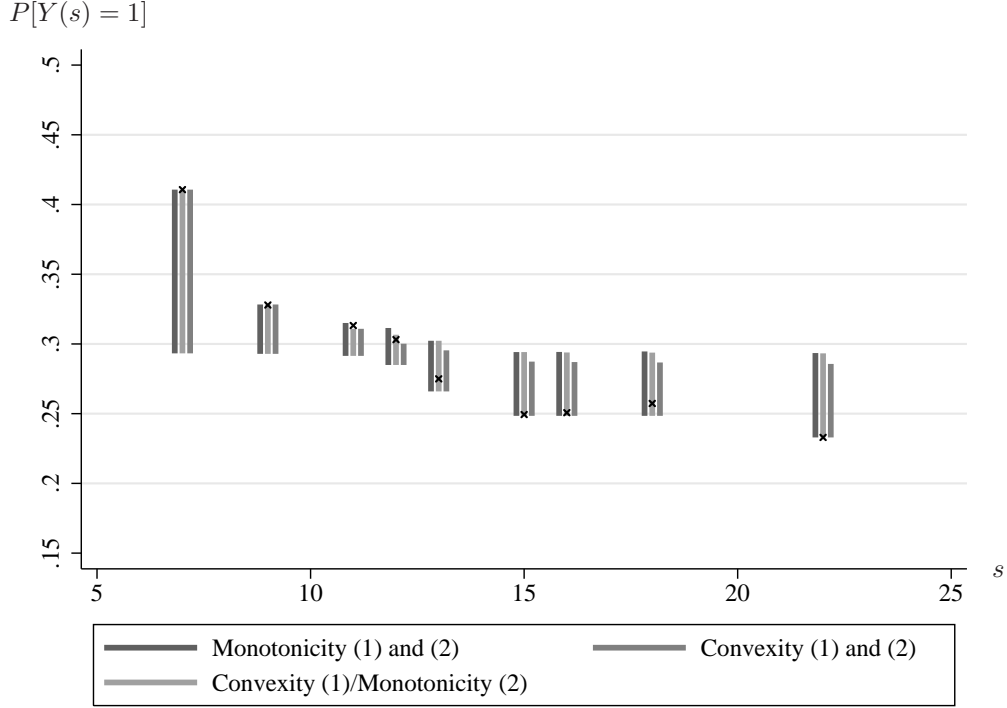
Figure 2: Bounds on the potential propensity to smoke
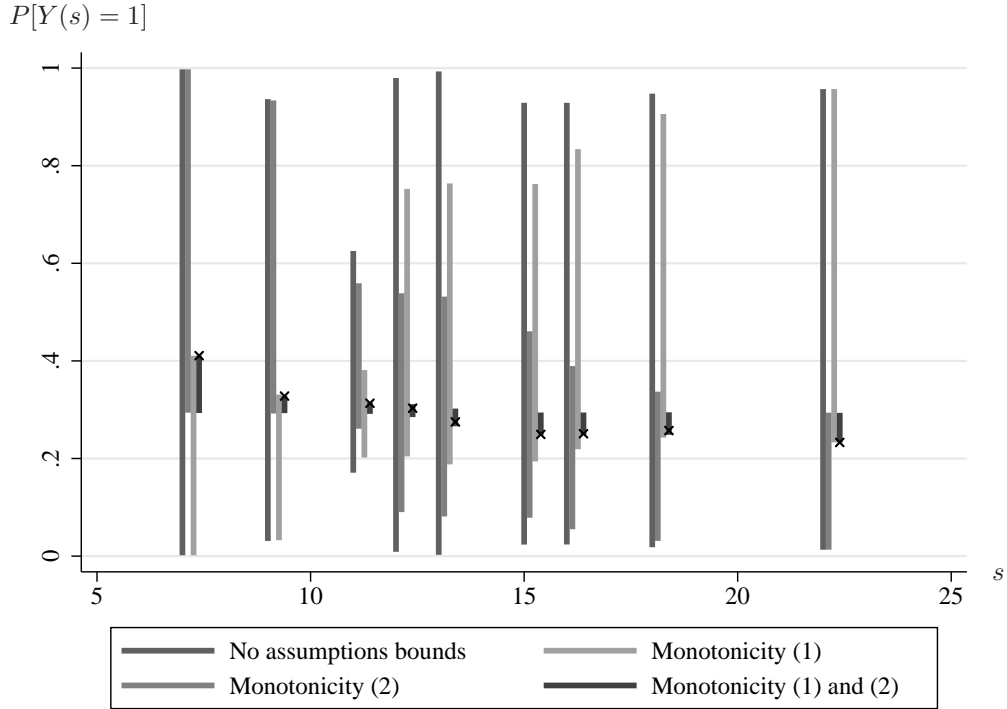
Source: Swiss Health Survey 2007, own calculations. *Notes:* $P[Y(s) = 1]$ denotes the potential propensity to smoke (Y) if everybody in the population was assigned s. Estimated bounds are regression adjusted for the respondents' age and gender. Bars indicate the width of the bounds. Left bars show the upper/lower bounds under no assumptions, middle bars additionally impose monotonicity in the sense of (1) or (2), right bars show the bounds under monotonicity (1) and (2). Crosses indicate the propensity to smoke if schooling was assumed exogenous (which correspond to the observed data points in Figure 1).
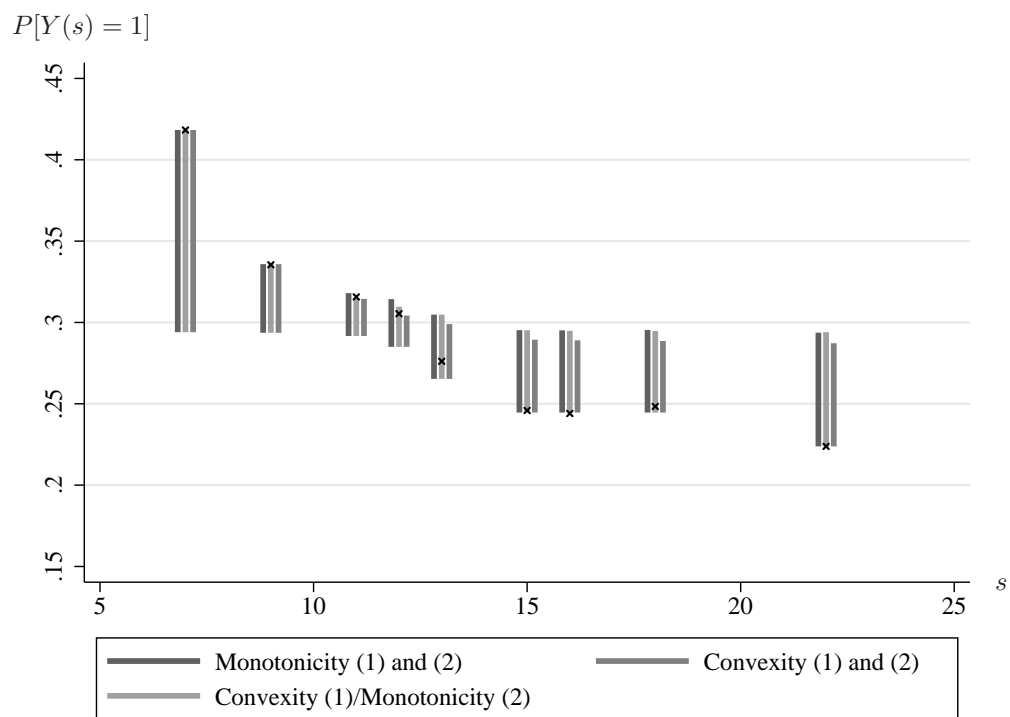
Figure 3: Bounds on the potential propensity to smoke

Figure 4: Bounds on the potential propensity to smoke with income fixed

Source: Swiss Health Survey 2007, own calculations. *Notes:* See Figure 3. Estimated bounds are regression adjusted for the respondents' age, gender, and income.

Table 2: Lower bounds on the causal effect of schooling on smoking

Schooling $s_1 \rightarrow s_2$	Observed difference	No ass.	Monotonicity (1) and (2)	Convexity (1) and Monotonicity (2)	Convexity (1) and (2)
$7 \rightarrow 9$	-0.0827	-0.9663	-0.1179	-0.1179	-0.1179
		(0.0017)	(0.0723)	(0.0724)	(0.0723)
$9 \rightarrow 11$	-0.0147	-0.7654	-0.0370	-0.0370	-0.0370
		(0.0046)	(0.0160)	(0.0160)	(0.0160)
$11 \rightarrow 12$	-0.0151	-0.6166	-0.0335	-0.0293	-0.0293
		(0.0053)	(0.0195)	(0.0203)	(0.0203)
$12 \rightarrow 13$	-0.0231	-0.9768	-0.0437	-0.0405	-0.0341
		(0.0015)	(0.0308)	(0.0197)	(0.0106)
$13 \rightarrow 15$	-0.0256	-0.9695	-0.0537	-0.0537	-0.0470
		(0.0018)	(0.0204)	(0.0205)	(0.0312)
$15 \rightarrow 16$	0.0013	-0.9051	-0.0456	-0.0456	-0.0389
		(0.0028)	(0.0099)	(0.0099)	(0.0173)
$16 \rightarrow 18$	0.0066	-0.9106	-0.0456	-0.0453	-0.0386
		(0.0029)	(0.0099)	(0.0101)	(0.0175)
$18 \rightarrow 22$	-0.0244	-0.9345	-0.0615	-0.0607	-0.0538
		(0.0027)	(0.0199)	(0.0195)	(0.0233)

Source: Swiss Health Survey 2007, own calculations. *Notes:* Estimated lower bounds on the causal effect $P[Y(s_2) = 1] - P[Y(s_1) = 1]$, where Y denotes the smoking indicator. Column 2 shows $P(Y = 1|S = s_2) - P(Y = 1|S = s_1)$. The following columns impose assumptions with increasing strength; column 3 under no assumptions, column 4 under joint monotonicity (1, ii) and (2, ii), column 5 additional convexity (1, i), column 6 under convexity (1) and (2). Estimates are regression adjusted for the respondents' age and gender. By assumption, the upper bound on the causal effect is zero under the monotonicity/convexity assumptions. Bootstrapped standard errors in parentheses.

Appendix

Table 3: Schooling categories and assigned years of schooling

Category	Years of schooling
Early drop out of compulsory school	7
Compulsory school	9
Basic vocational education and training	11
Matura schools	12
Second vocational training	13
Higher vocational education and training	15
University of applied science; university of teacher education	16
University including federal institutes of technology	18
PhD/Doctorate	22

Notes: Highest education achieved. For an overview of the Swiss education system, see http://www.edudoc.ch/static/web/bildungssystem/grafik_bildung_e.pdf (last accessed 2010, Oct 13).