Are estimated peer effects on smoking robust?: Evidence from adolescent students in Spain

R. Duarte

Department of Economic Analysis, University of Zaragoza, Gran Vía, 2. 50005. Zaragoza. Spain.
e-mail: *rduarte@unizar.es*

J-J. Escario (corresponding author)

Department of Economic Analysis, University of Zaragoza, Gran Vía, 2. 50005. Zaragoza. Spain. Phone: 34 974 239373, Fax: 34 974 239375.
e-mail: *jescario@unizar.es*

J-A. Molina

Department of Economic Analysis, University of Zaragoza, Gran Vía, 2. 50005. Zaragoza. Spain. And Institute for the Study of Labor-IZA (Bonn, Germany)
e-mail: *jamolina@unizar.es*

Abstract: Adolescent smoking is one of the most pressing public health problems. The objective of this paper is to analyse the influence of peer pressure on adolescent cigarette consumption. More concretely, we explore the significance and robustness of the peer effects using several estimation methods employed in the existing literature. On the basis of the data provided by the 2004 Spanish Survey on Drug Use in the School Population, we estimate the probability of being a smoker by two-stage models. The results reveal that when we use standard errors used in the literature the class peer variable appears to be significant. However, the class peer variable is not significant when we calculate more exigent standard errors, a result that is robust across all specifications. The paper suggests the need for a more cautious interpretation of the peer effects found previously in the literature until a deeper analysis confirms the robustness of the peer effects.

Keywords: Peer effects, tobacco consumption, adolescents, bootstrap

JEL classification: D71, I10, I12, I18

1. Introduction

Tobacco consumption among adolescents is one of the most pressing public health problems, with youth smoking being recognized as the major preventable risk factor for cancer. Most smokers initiate this behaviour in adolescence (Yu and Williford 1992; Smith et al. 2007), and the early starting age doubles the risk of lung and other types of cancer (Taylor 1993; Gold et al. 1996; Band et al. 2002). Moreover, the consumption of tobacco appears to be addictive, most young smokers begin smoking with no intention of becoming smokers, but they are trapped by the addiction. Thus, Becker and Murphy (1988) develop a model in which past consumption increases the marginal utility of present consumption.

Despite these facts, tobacco is the most widely consumed drug among adolescents, with a number of aspects of this consumption being studied in recent years, e.g. the impact of addiction (Chaulopka 1991; Becker et al. 1994; Pierani and Tiezzi 2009), the impact of taxes (Escario and Molina 2004; Wan 2006; Tauras et al. 2007), and the effects of advertising (Keeler et al. 2004).

Looking at the explanatory factors of tobacco consumption among adolescents, the immediate and broader social environment of the individual, that is to say, family, friends, and school community, appear to be important influences on the initiation of this risk behaviour (Flay et al. 1999; Leatherdale et. al. 2006; Lundborg 2006; Ali and Dwyer 2009). Papers such as Mounts and Steimberg (1995), Harturp (1999) and Dekovic et. al. (2004) explore these factors, arguing that, during adolescence, the group of friends and schoolmates becomes the most important social reference for students.

The literature has often focused on these social environments (immediate and broader), studying the correlates between drug consumption and peer behaviour

(Gaviria and Raphael 2001; Eisenberg and Foster 2003; Scal et al. 2003; Powel et al. 2005; Lundborg 2006; Clark and Lohéac 2007; Ali and Dwyer 2009; Duarte et al. 2011; McVicar 2011). In most cases, positive and significant correlates are found between individual smoking behaviour and the behaviour of peers. An excellent overview about theorical and empirical literature on peer effects can be found in Epple and Romano (2011), although the survey is focused in the education topic.

Manski (1993) warned of the difficulty of inferring whether or not the average behaviour in a certain group influences the behaviour of the individuals comprising that group, due to the "reflection" problem. Other difficulties arise from the existence of different types of social interaction. Following Manski (1993), three hypotheses explain why individuals who belong to the same group tend to behave similarly. First, the propensity of an individual to behave in a certain way is influenced by the behaviour of the group (*endogenous effects*). The second hypothesis, *exogenous or contextual effects*, establishes that the propensity of an individual to behave in a certain way varies with the exogenous characteristics of the peer group. Finally, the third level of social interaction, or *correlated effects,* arises from the fact that individuals in the same group tend to behave similarly, since they share similar characteristics and institutional environments. As Manski (1993) showed, if we assume the existence of both contextual and endogenous effects, there is no clear way to distinguish between them. Consequently, most papers have assumed that there are no contextual effects.

An important issue pointed out by Foster (2006), in the educational context, is that most peer effects found in education are modest and not robust to alternative specifications, and as a consequence it is necessary to be more cautious in interpreting the results. Following this idea, our objective is to analyse the robustness of the peer effect significance in the context of smoking behaviour. In particular, we investigate

whether this peer effect is robust when we take into account the possible correlation in the unobserved variables of the same group of peers, in order to compute standard errors.

On the basis of the data provided by the Spanish Survey on Drug Use in the School Population, collected for 2004 by the Spanish Government's Delegation for the National Plan on Drugs, we compute a peer group measure at the class level, and analyse its impact on the decision of adolescents to smoke.

The rest of the paper is structured as follows. Section II is devoted to presenting the model, as well as the empirical strategy. The data and variables are described in Section III. In Section IV, the results are presented, and Section V closes the paper with a review of our main conclusions.

2. Model and strategy

In order to examine the determinants of adolescent tobacco consumption, we assume that people make rational choices, in the sense that they decide to smoke if the utility of such behaviour is higher than the utility of not smoking. Accordingly, each student i belonging to class c in school k, indexed by ick, has a utility function $u_{ick}(Y_{ick}, X_{ick}, P_{ick})$, where variable Y_{ick} is a dichotomous variable that takes values one or zero, depending on whether the adolescent smokes or not, X_{ick} is a set of covariates reflecting the individual and family characteristics of the adolescent, and P_{ick} is the peer group variable, measuring the prevalence of cigarette consumption in the reference group of the adolescent. The individual will decide to smoke cigarettes if the following condition holds:

$$u_{ick}(1, X_{ick}, P_{ick}) - u_{ick}(0, X_{ick}, P_{ick}) > 0 \qquad (1)$$

while he/she will prefer to be a non-smoker otherwise.

Given that the dependent variable is the outcome that reflects the individual decision, most empirical research has been carried out within the random utility framework used in binary choice models, adopting the following stochastic specification:

$$Y_{ick}^* = u_{ick}(1, X_{ick}, P_{ick}) - u_{ick}(0, X_{ick}, P_{ick}) = \beta X_{ick} + \gamma P_{ick} + \varepsilon_{ick} \qquad (2)$$

where Y_{ick}^* is a latent variable whose sign determines whether or not the individual consumes tobacco, that is to say, the value (one or zero) of the observable variable Y_{ick}. Finally, ε_{ick} is an error term normally distributed with mean zero and unitary variance.

Despite the fact that, until Case and Katz (1991), empirical work used peer effect regressors as exogenous variables, the estimation of equation (2) must account for some potential sources of endogeneity. First, the individual error would be correlated with the peer effect variable if we assume, which is the most plausible option, that individual smoking affects the average smoking of the group of friends. Moreover, adolescents who become smokers will tend to have friends who smoke. In other words, we are assuming that there is causality in both directions. Thus, average smoking affects the individual decision of whether or not to smoke and, at the same time, individual behaviour also affects the behaviour of the group of friends. Second, families could sort across schools according to their conscientiousness about the education of their children, in such a way that they devote resources to the choice of where to live, focusing on the perceived quality of schools in the area. Both of these endogeneity problems would yield upwardly biased peer effect estimates.

As we have pointed out above, it was not until the 1990s that research took into account the possible endogeneity of the peer variable, and the majority of subsequent studies have implemented an instrumental approach, as do we. In this context, models to

analyse the decision to smoke or not, have often been estimated by a two-stage procedure. Following this approach, we estimate equation 2 using a Probit model[1].

Following Gaviria and Raphael (2001), Krauth (2007), and Lundborg (2006), among others, we assume there are no contextual effects and thus parameter γ can be interpreted as the endogenous social effect or endogenous peer effect[2]. Further, this assumption enables us to solve the problem of bidirectional causality, as the individual behaviour is not affected by the average characteristics of peers, consequently, they are natural candidates to instrument the peer variable.

With respect to the problem of sorting, prior studies have found that this problem is not very significant (Gaviria and Raphael 2001; Lundborg 2006). In particular, Gaviria and Raphael (2001) split the sample into two groups: "movers", those youths whose families moved during the last two years, and "stayers", the remainder. The hypothesis is that, under relevant endogenous sorting, the peer effects will be larger for "movers". They find no evidence of statistically significant differences in the peer effect estimates for four of the five behaviours analysed, with one of these being cigarette smoking. Similarly, Lundborg (2006) includes school/grade fixed effects in order to address the issue of sorting. After estimating the peer effects with and without school/grade fixed effects, the magnitude of the peer effects in smoking behaviour decreases only 7%.

---

[1] In general, the literature that estimated Probit models reported the standard errors of the second stage with the robust option, but without correcting for the generated peer regressor (see, for example, Lundborg (2006) or Clark and Lohéac (2007)).

[2] As Krauth (2007) pointed out, in most cases it is only possible to distinguish between endogenous and contextual effects by assuming that one or the other is absent. Thus, as Lundborg (2006) pointed out, by using measures of peer effects at the class-level, the importance of the contextual effects will be reduced, since when the reference group is broader, pupils are likely less exposed to the family background of their peers, and thus the observed peer effects are more likely to be caused by endogenous rather than contextual effects.

The sorting problem is probably less important in Spain than in other countries, given that parents cannot choose which class their sons and daughters attend. In fact, students are assigned to different classes in alphabetical order or by other school rules. Moreover, most students in Spain are sent to the schools of the district where the youth lives; there is a set of norms to assign a particular school, and living in the district of the school is the most important criterion. However, some sorting is still possible, that is, parents can move to another school district, or they can choose to send their child to a private school, with this second type of sorting being the most significant.

Consequently, in order to deal with these endogeneity and sorting problems, we implement the following strategy. First, our estimations include school fixed effects. Second, given that the sorting reflects the unobservable "preoccupation" of parents, the introduction of several variables that measure parents' backgrounds, such as education level, will reduce spurious correlations, as these variables are "good" proxies for the interest of parents in giving their sons and daughters a good education. Finally, we use instrumental variables to deal with any other correlations between the peer variables and the error term. With these strategies, although some unobserved heterogeneity could still remain, we consider that the problem is considerably alleviated.

3. Data and variables

In order to implement the empirical model, the data analyzed in this paper come from the Spanish Survey on Drug Use in the School Population for the year 2004. This survey is carried out by the Spanish Government's Delegation for the National Plan on Drugs, and constitutes a nationally representative sample of the student population between 14 and 18 years old. We restrict our sample to the years of compulsory education, up to age 16, since the population between 16 and 18 years old are not

randomly sampled and those who decide to drop out of school cannot be interviewed[3]. The information provides measures of smoking behaviour and certain individual and family characteristics. The questionnaires were filled in confidentially, in a classroom setting. The survey was carried out in both state/public and private centres of secondary education and vocational training throughout the national territory.

The dependent variable in the study is *CigaretteConsumption*, a dichotomous variable which takes value 1 if the individual has smoked cigarettes during the last 30 days, and 0 otherwise. In our data, most occasional smokers are not considered as smokers, since those who have declared having smoked only a few times in their lives are asked to skip the following questions about smoking behaviour, among which is the question whether or not they have smoked cigarettes during the last 30 days. Consequently, we consider those individuals to be non-smokers.

Given that the main objective of this work is to evaluate the peer effect on individual cigarette consumption, we must define an appropriate measure of the peer group. As we have mentioned, we use a traditional measure of the peer effect computed at class level. For an individual i belonging to class c in school k, peer group cigarette consumption is defined by taking the class average prevalence of tobacco consumption, after subtracting his/her contribution to this average:

---

[3] Thus, the final sample size is 10,666 individuals, of which 10,102 can be used in our estimations. According to our data set, first smoking decisions are mostly taken by Spanish adolescents before age 16. Specifically, 23.13 % of respondents aged 15 are smokers. Although it is difficult to generalize the results of the paper to broader age groups, we feel we are justified in doing so, since i) this figure of 23.13% of smokers among people who are 15 years old is a significant proportion of future smokers (37% male, 28% female according to the 2008 WHO report), ii) the fact that we are using a large sample with more than 10,000 students, and iii) the fact that the estimated marginal peer effects (0.558) are in the normal range found in the literature. Nevertheless, a caution must be introduced regarding the different gender composition of the adolescent and adult populations.

$$P_{ick} = \frac{1}{N_{ck} - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N_{ck}} Y_{jck}$$

where N_{ck} is the total number of individuals in this class.

Apart from the peer variable, other exogenous variables are considered, including individual and family characteristics. Additionally, as mentioned in the previous section, school fixed effects are considered in our analysis. The definition and a descriptive analysis of these variables are shown in Table I.

(Table 1)

4. Empirical results

In order to deal with the endogeneity problem of the peer effects variables, we use a two-stage procedure that provides consistent estimates of equations 2 and 3. We cannot use the estimated variances of the second stage, as they ignore the fact that a regressor has been estimated. One general approach, applicable to both the linear probability model and the Probit model, is to use bootstrap methods for statistical inference. As Cameron and Trivedi (2005, p. 355) point out, a bootstrap procedure can lead to a more refined asymptotic theory. In this application, we use normal bootstrap and block-bootstrap procedures.

With respect to the first stage of the estimation, we have regressed the peer class-based variable on the exogenous variables and five instruments that are the class average, after excluding individual i of the following variables: *Income, SmokingMother, SmokingFather, WithoutFather* and *HouseWifeMother*.

In order to gain confidence in the instruments used, we have implemented several tests. We test the joint significance of the instruments[4] and the over-identification restrictions by following Wooldridge (2002, p. 122-123). The results provide evidence that the instruments are clearly significant in explaining the peer effect variable, after controlling for the exogenous covariates (F=135.82, p-value= 0.000), and thus we cannot reject the null hypothesis of the validity of the instruments (W=7.085, with the critical value being $\chi^2_{.95}(5) = 11.1$).

Having explained the procedure to instrument the peer effect, we go now to obtain the estimates of equations 2 and 3, beginning with Equation 2, which is estimated by a Probit model. To that end, we use two strategies. In the first, we follow the two-step approach suggested by Rivers and Vuong (1988), using the original peer effect variable, and we introduce the first stage errors as an additional regressor (see Table 2). In the second, we replace the peer effect variable by its predicted counterpart obtained in the first stage linear regression, reported in Table A.1. The estimates appear in Table 3.

In both cases, we present three different estimates of the t-statistics for the coefficient estimates. In the first column we do not correct the standard errors of the second stage in order to compute the t-ratios, in the second column we correct the standard errors with normal bootstrap and, finally, in the third column, the standard errors have been corrected with block-bootstrap. In both cases, normal bootstrap and block-bootstrap procedures use 200 replications (Efron and Tibsharani (1993) consider that this practice is almost always enough for standard error estimation).

(Table 2)

---

[4] The first-stage estimates are available upon request.

The computation of standard errors taking into account the cluster structure of the data is important in order to account for the third level of social interaction, or *correlated* effects, that are independent of the exogenous variables. Thus, students in the same class are taught by the same teacher, who can exert heavy, little or even no effect on their smoking behaviour. Hence, individuals in the same class tend to behave similarly since they share similar institutional environments. Although omitting these correlated effects does not yield inconsistent estimates of the coefficients if they are independent, it is necessary to take them into account for efficiency reasons, and for the computation of correct standard errors.

The two-stage procedure of estimation following Rivers and Vuong (1988) enables us to obtain marginal effects (see Table 2). In order to compute average partial effects or marginal effects under the Rivers and Vuong approach, we follow Wooldridge (2002), using his equation (15.47), $\frac{1}{N}\sum_{i=1}^{N}\Phi(\hat{\beta}_{\rho_1}X_{ck}+\hat{\gamma}_{\rho_1}P_{ck}+\hat{\theta}_{\rho_1}\hat{\upsilon}_{ick})$, and, for continuous variables, we take derivatives, or, for discrete variables, we evaluate the difference of the equation at two values of the variable. For example, for a continuous variable x_j, the average partial effect would be $\frac{1}{N}\sum_{i=1}^{N}\phi(\hat{\beta}_{\rho_1}X_{ck}+\hat{\gamma}_{\rho_1}P_{ck}+\hat{\theta}_{\rho_1}\hat{\upsilon}_{ick})\beta_j$, where the $\hat{\upsilon}_{ick}$ are the first-stage OLS residuals and $\hat{\beta}_{\rho_1}, \hat{\gamma}_{\rho_1}$ and $\hat{\theta}_{\rho_1}$ are the estimated coefficients in the Probit second stage, where the first stage errors have been introduced as an additional variable. The marginal effects for the residuals of the first stage, and for the *IncomeSquared* variable, are not computed as it has no economic interpretation for the residuals and we have computed the marginal effect of disposable income taking into account the effects of both *Income* and *IncomeSquared* variables.

The peer effect estimates are listed at the top of Table 2. A noticeable result is that the peer effect at the class level is significant in two of the specifications. With

reference to the quantitative impact of the peer effect, it appears to be quite high, although it is in line with most of the literature (see Krauth 2007, p. 296, for a survey of the quantitative estimates found in several studies). According to Table 2, our estimates suggest that if students are in a class where the percentage of smokers is 10 points higher, all other things being equal, their probability of becoming smokers will be 5.58 points higher.

A different issue is the significance of the results. Thus, the peer effect is statistically significant at the 5% level when we do not correct standard errors, or when we correct them by normal bootstrapping. However, only without correction does the peer effect appear significant at the 1% level. Moreover, the t-statistic with the normal bootstrap correction is barely higher than 2, which is a modest value if we take into account that the estimate uses a considerable sample of 10,102 adolescents. A different result appears if we use a block-bootstrap procedure, in which case the peer effect is not significant, even at the 10% level.

This result is in line with Krauth (2007) who estimates a structural model which takes non-random peer selection into account, by considering within-group correlation in the unobservable variables or error terms. The model uses this correlation in the estimation of both coefficients and standard errors and, as in our case, finds t statistics for the peer effect below the usual asymptotic critical values.

All in all, the results of this two-stage estimation yield mixed evidence about the importance of the peer effect in smoking behaviour among adolescents. The reported peer effect is quantitatively high. However, the significance is not robust to alternative specifications, and when we use the more refined standard errors to take into account the cluster nature of the data, the peer effect is clearly not significant.

In relation to the second two-stage procedure (see estimates in Table 3), it is important to point out that the fitted values approach, a common procedure in the literature, does not allow us to obtain consistent average partial effects (see for example, Wooldrige 2002; or Arellano 2008). As a consequence, we report the estimated coefficients that, as they are a scale transformation of the structural parameters, only permit the determination of the direction of the effects and, therefore, whether these are significant or not. Although, the t-statistics are somewhat higher, our main conclusion remains. That is to say, the coefficient of the peer variable (or the direction of the effect) is positively significant with no correction or normal bootstrap correction of the standard errors, but it appears not to be significant when the cluster nature of the data is taken into account[5].

(Table 3)

5. SUMMARY AND CONCLUSIONS

The objective of this paper was to go deeper into the analysis of peer effects among adolescents who face the decision of whether or not smoke. More concretely, we analyse the robustness of the peer effects using several estimation methods employed in the existing literature. To that end, we compute several standard errors for every estimation method.

---

[5] In order to check the stability and, therefore, the robustness of these results, we estimate a linear probability model with the same three versions of the standard errors used in the Probit models. Once again, our results provide evidence that the peer effect is not robust when more restrictive standard errors, accounting for the cluster nature of the data, are considered. It should be noted that this conclusion is robust to several specifications, that is to say, in the linear probability model and in the Probit model with two estimation approaches. The estimates of the linear probability model are available upon request.

We have examined the determinants of being a smoker by estimating a Probit model and a probability linear model, using the data provided by the Spanish Survey on Drug Use in the School Population (2004).

Although we have found positive and significant peer effects for some specifications, we find that social interactions or peer effects are not significant once we calculate more "refined" or exigent standard errors. It appears that block bootstrap can yield more appropriate standard errors as it takes into account the nature of the clustered sample. Moreover, even if we do not favour either of both bootstrap standard errors, the results provide evidence that the peer effects are not robust to different specifications of the standard errors. The plausibility of the results is reinforced by the similarity of the results across all specifications.

The paper also highlights the difficulties of achieving conclusive results about peer effects (McVicar, 2011), a problem first pointed out by Manski (1993). The relationship between the individual decision and the peer variable is increasing in γ (the endogenous effect) and in the correlation of the unobserved variables of the peer group (correlated effects). What our paper illustrates is that, if we consider this correlation when computing standard errors, which appears as an advisable strategy, the peer effects are no longer significant and, consequently, doubts could arise about the existence of endogenous peer effects commonly found in the literature about smoking.

The existence, or not, of peer effects, acting as "social multipliers", has important policy implications. Policy interventions in preventing tobacco consumption among adolescents could exert larger gains in terms of reducing smoking rates, since effective policies act at the individual level, reducing the number of adolescent smokers. However, if the number of individuals who smoke declines, then the fraction of peers who smoke also declines and, consequently, this will have an indirect effect in reducing

the number of adolescent smokers as they have fewer smoker friends and there would then be less pressure to smoke. These indirect or amplifying effects are known as social multipliers.

We conclude that our findings should encourage more caution in the interpretation of peer smoking effects found in most of the literature on peer effects, and foster more robust analysis in the future. Thus, although marginal effects at the class level are high, their statistical significance can be not robust to alternative procedures of computing standard errors, which gives rise to certain doubts about the social multipliers of policy interventions. As a consequence, future research needs to analyse more thoroughly the robustness of peer effects under alternative specifications.

References

Ali, M, Dwyer D (2009).Estimating peer effects in adolescent smoking behaviour: a longitudinal analysis. J Adolescent Health 45: 402-408

Arellano M (2008) Binary Models with Endogenous Explanatory Variables. http://www.cemfi.es/~arellano/binary-endogeneity.pdf. Accessed 16 January 2012

Band PR, Le ND, Fang R, Deschamps M (2002) Carcinogenic and endocrine disrupting effects of cigarette smoke and the risk of breast cancer. Lancet 360: 1044-1049

Becker GS, Murphy KM (1988) A theory of rational addiction. J Polit Econ 96: 675-700

Becker GS, Grossman M, Murphy KM (1994) An empirical analysis of cigarette addiction. Am Econ Rev 84: 396-418

Cameron AC, Trivedi PK (2005) Microeconometrics: Methods and Applications. Cambridge University Press, Cambridge

Case AC, Kartz LF (1991) The Company You Keep: The Effects of Family and Neighbourhood on Disadvantaged Youths. NBER Working Paper No. 3705

Chaloupka F (1991) Rational addictive behavior and cigarette smoking. J Polit Econ 99: 722-742

Clark AE, Lohéac Y (2007) It wasn't me, it was them! Social influence in risky behaviour by adolescents. J Health Econ 26: 763-784

Dekovic M, Wissink IB, Meijer AM (2004) The role of family and peer relations in adolescent antisocial behaviour: comparison of four ethnic groups. J Adolescence 27: 497-514

Duarte R, Escario J, Molina J (2011) Peer effects, unobserved factors and risk behaviours in adolescence. Rev Econ Aplicada, 55: 125-151

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, London

Eisenberg ME, Foster JL (2003) Adolescent smoking behaviour: measures of social norms. Am J Prev Med 25: 122-128

Epple D, Romano RE (2011) Peer effects in Education: A survey of the theory and evidence. Benhabib J, Bisin A, Jackson M (ed) Handbook of Social Economics 1, Elsevier Sciencie, North Holland, pp 1053-1163

Escario JJ, Molina JA (2004) Will a special tax on tobacco reduce lung cancer mortality? Evidence for EU countries. Appl Econ 36: 1717-1722

Flay BR, Petraitis J, Hu FB (1999) Psychosocial risk and protective factors for adolescent tobacco use. Nicotine Tob Res 1: S59-S65

Foster G (2006) It's not your peers, and it's not your friends: some progress toward understanding the educational peer effect mechanism. J Public Econ 90: 1455-1475

Gaviria A, Raphael S (2001) School-based peer effects and juvenile behaviour. Rev of Econ Stat 83: 257-268

Gold Dr Wang X, Speizer FE, Ware JH, Dockery DW (1996) Effects of cigarette smoking on lung function in adolescent boys and girls. New Engl J Med 335: 931-937

Harturp WW (1999) Constraints on peer socialization: Let me count the ways. Merrill Palmer Quart 45: 172-183

Keeler T, Hu T, Ong M, Sung H (2004) The US National tobacco settlement: the effects of advertising and price changes on cigarette consumption. Appl Econ 36: 1623-1629

Krauth B (2007) Peer and selection effects on youth smoking in California. J Bus Econ Stat 25: 288-298

Leatherdale ST, Cameron R, Brown KS, Jolinm MA, Kroeker C (2006) The influence of friends, family and older peers on smoking among elementary students: Low-risk students in high-risk schools. Prev Med 42: 218-222

McVicar D (2011) Estimates of peer effects in adolescent smoking across twenty six European Countries. Soc Sci Med 73: 1186-1193

Lundborg P (2006) Having the wrong friends? Peer effects in adolescent substance use. J Health Econ 25: 214-233

Manski CF (1993) Identification of endogenous social effects: The refection problem. Rev Econ Stud 60: 531-542

Mounts NS, Steimberg L (1995) An ecological analysis of peer influence on adolescent grade point average and drug use. Dev Psychol 31: 915-922

Pierani P, Tiezzi S (2009) Addiction and interaction between alcohol and tobacco consumption. Empirical Economics 37: 1-23

Powell LM, Tauras JA, Ross H (2005) The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behaviour. J Health Econ 24: 950-968

Rivers D, Vuong QH (1988) Limited information estimators and exogeneity tests for simultaneous probit models. J Econometrics 39: 347-366

Scal P, Ireland M, Borowsky IW (2003) Smoking among American adolescents: a risk and protective factor analysis. J Commun Health 28: 79-97

Smith BJ, Phongsavan P, Bauman AE, Havea D, Chey T (2007) Comparison of tobacco, alcohol and illegal drug usage among school students in three Pacific Island societies. Drug Alcohol Depen 88: 9-18

Spanish Government's Delegation for the National Plan on Drugs (2004) Spanish Survey on Drug Use in the School Population

Tauras J, Powell L, Chaloupka F, Ross H (2007) The demand for smokeless tobacco among male high school students in the United States: the impact of taxes, prices and policies. Appl Econ 39: 31-41

Taylor R (1993) Risk of premature death from smoking in 15-year-old Australians. Aust J Public Health 17: 358-364

Wan J (2006) Cigarette tax revenues and tobacco control in Japan. Appl Econ 38: 1663-1675

Wooldridge, JM (2002) Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press

Yu J, Williford WR (1992) The age of alcohol onset on alcohol, cigarette, and marijuana use patterns: an analysis of drug use progression of young adults in New York State. Int J Addict 27: 1313-1323

Table 1. Descriptive analysis

Variable	Definition	Mean (Std. Deviation)
TobaccoConsumption	This takes value 1 if the adolescent has smoked more than one cigarette per day in the last month and 0 otherwise	0.198 (0.398)
ClassPeerGroup	Smoking prevalence in the class after eliminating the individual's influence	0.236 (0.136)
Gender	This takes the value 1 if the young person is male and 0 if female	0.488 (0.500)
Age14	This takes value 1 if the adolescent is 14 years old and 0 otherwise (*omitted category*)	0.339 (0.473)
Age15	This takes value 1 if the adolescent is 15 years old and 0 otherwise	0.661 (0.473)
WithoutFather	This takes value 1 if the adolescent lives without the father at home and 0 otherwise	0.112 (0.315)
UnemployedFather	This takes value 1 if the father of the adolescent is unemployed and 0 otherwise	0.014 (0.116)
HouseWifeMother	This takes value 1 if the mother is a housewife and 0 otherwise	0.320 (0.467)
SmokerFather	This takes value 1 if the father of the adolescent smokes and 0 otherwise	0.324 (0.468)
SmokerMother	This takes value 1 if the mother of the adolescent smokes and 0 otherwise	0.325 (0.468)
FatherDrinking	This takes the value 1 if he/she never drinks alcohol; 2 if he/she sometimes drinks; 3 only weekends; 4 almost everyday in moderation; 5 everyday a lot.	2.236 (1.259)
MotherDrinking	This takes the value 1 if he/she never drinks alcohol; 2 if he/she sometimes drinks; 3 only weekends; 4 almost everyday in moderation; 5 everyday a lot.	1.754 (1.050)
PrimaryStudiesMother	This takes value 1 if the mother has no more than basic school certificate and 0 otherwise (*omitted category*)	0.507 (0.500)
SecondaryStudiesMother	This takes value 1 if the mother has a secondary school certificate or vocational training but not a university degree and 0 otherwise	0.253 (0.434)
UniversityStudiesMother	This takes value 1 if the mother has a university degree and 0 otherwise	0.240 (0.427)
PrimaryStudiesFather	This takes value 1 if the father has no more than basic school certificate and 0 otherwise (*omitted category*)	0.500 (0.500)
SecondaryStudiesFather	This takes value 1 if the father has a secondary school certificate or vocational training but not a university degree and 0 otherwise	0.276 (0.447)
UniversityStudiesFather	This takes value 1 if the father has a university degree and 0 otherwise	0.224 (0.417)
Income	Available income per week of the adolescent (in euros)	12.901 (14.517)

Table 2. Estimation results Probit model (Rivers and Vuong Marginal effects)

Variable	Without correction		Normal bootstrap		Cluster bootstrap	
ClassPeerGroup	0.558	***	0.558	**	0.558	
	(2.588)		(2.045)		(0.642)	
Residual(First-Stage)[a]	-		-		-	
	-		-		-	
Gender	-0.083	***	-0.083	***	-0.083	***
	(-10.439)		(-9.305)		(-8.110)	
Age15	0.103	***	0.103	***	0.103	***
	(10.541)		(8.960)		(4.273)	
WithoutFather	0.054	***	0.054	***	0.054	***
	(3.334)		(3.523)		(3.046)	
UnemployedFather	0.057		0.057		0.057	
	(1.464)		(1.373)		(1.300)	
HouseWifeMother	-0.019	**	-0.019	**	-0.019	*
	(-2.254)		(-2.008)		(-1.825)	
SmokerFather	0.031	***	0.031	***	0.031	***
	(3.475)		(3.628)		(3.573)	
SmokerMother	0.047	***	0.047	***	0.047	***
	(5.243)		(5.019)		(5.152)	
FatherDrinking	0.003		0.003		0.003	
	(0.758)		(0.731)		(0.681)	
MotherDrinking	-0.007		-0.007		-0.007	
	(-1.603)		(-1.462)		(-1.407)	
SecondaryStudiesFather	-0.016		-0.016		-0.016	
	(-1.467)		(-1.492)		(-1.289)	
UniversityStudiesFather	-0.009		-0.009		-0.009	
	(-0.703)		(-0.637)		(-0.629)	
SecondaryStudiesMother	-0.010		-0.010		-0.010	
	(-0.917)		(-0.863)		(-0.777)	
UniversityStudiesMother	-0.023	*	-0.023	*	-0.023	
	(-1.891)		(-1.743)		(-1.546)	
Income	0.006	***	0.006	***	0.006	***
	(11.141)		(7.444)		(7.790)	
IncomeSquared[b]	-		-		v	
	-		-			
Nº Observations	10102		10102		10102	

Notes: t-statistics within parentheses. * significant at the 10% level. ** significant at the 5% level. *** significant at the 1% level.

[a,b] Marginal effects are not computed because they have not economic interpretation

Table 3. Estimation results Probit model (Coefficients using fitted values)

Variable	Without correction		Normal bootstrap		Cluster bootstrap	
ClassPeerGroup	2.235	***	2.235	***	2.235	
	(2.525)		(2.485)		(1.360)	
Gender	-0.341	***	-0.341	***	-0.341	***
	(-10.232)		(-8.997)		(-8.207)	
Age15	0.455	***	0.455	***	0.455	***
	(9.984)		(8.793)		(6.991)	
WithoutFather	0.211	***	0.211	***	0.211	***
	(3.589)		(3.605)		(3.229)	
UnemployedFather	0.205		0.205		0.205	
	(1.493)		(1.360)		(1.274)	
HouseWifeMother	-0.082	**	-0.082	**	-0.082	*
	(-2.276)		(-2.147)		(-1.944)	
SmokerFather	0.127	***	0.127	***	0.127	***
	(3.562)		(3.619)		(3.671)	
SmokerMother	0.185	***	0.185	***	0.185	***
	(5.316)		(5.190)		(5.051)	
FatherDrinking	0.013		0.013		0.013	
	(0.840)		(0.746)		(0.766)	
MotherDrinking	-0.028		-0.028		-0.028	
	(-1.529)		(-1.289)		(-1.330)	
SecondaryStudiesFather	-0.071		-0.071		-0.071	
	(-1.572)		(-1.539)		(-1.470)	
UniversityStudiesFather	-0.038		-0.038		-0.038	
	(-0.729)		(-0.650)		(-0.691)	
SecondaryStudiesMother	-0.039		-0.039		-0.039	
	(-0.877)		(-0.830)		(-0.768)	
UniversityStudiesMother	-0.098	*	-0.098		-0.098	*
	(-1.851)		(-1.593)		(-1.660)	
Income	0.023	***	0.023	***	0.023	***
	(11.048)		(7.835)		(7.866)	
IncomeSquared	0.000	***	0.000	***	0.000	***
	(-6.682)		(-3.360)		(-3.331)	
Intercept	-1.569	***	-1.569		-1.569	***
	(-4.605)		(-1.323)		(-2.766)	
Nº Observations	10102		10102		10102	

Notes: t-statistics within parentheses. * significant at the 10% level. ** significant at the 5% level. *** significant at the 1% level.