Ruhr Economic Papers

Silja Göhlmann

The Determinants of Smoking Initiation

Empirical Evidence for Germany

#27

RWI
ESSEN

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstraße 150, 44801 Bochum, Germany

Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstraße 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen)
Hohenzollernstrasse 1/3, 45128 Essen, Germany

Editors:

Prof. Dr. Justus Haucap
RUB, Department of Economics
Competition Policy and Industrial Economics
Phone: +49 (0) 234/32 253 36, e-mail: justus.haucap@rub.de

Prof. Dr. Wolfgang Leininger
University of Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231 /7 55-32 97, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-36 55, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI Essen
Phone: +49 (0) 201/81 49-227, e-mail: schmidt@rwi-essen.de

Editorial Office:

Joachim Schmidt
RWI Essen, Phone: +49 (0) 201/81 49-292, e-mail: schmidtj@rwi-essen.de

Ruhr Economic Papers #27

Responsible Editor: Christoph M. Schmidt

ISSN 1864-4872 (online) – ISBN 978-3-86788-022-0

Ruhr Economic Papers
#27

Silja Göhlmann

The Determinants of Smoking Initiation

Empirical Evidence for Germany

RWI
ESSEN

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

ISSN 1864-4872 (online)
ISBN 978-3-86788-022-0

Silja Göhlmann*

The Determinants of Smoking Initiation – Empirical Evidence for Germany

Abstract
This paper aims at analyzing the determinants of the decision to start smoking using data from the German Socio-Economic Panel (GSOEP). The data used is a combination of retrospective information on the age individuals started smoking and, by tracing back these individuals within the panel structure up to the point they started smoking, information on characteristics at the age of smoking initiation. In contrast to other papers, it is possible to control for the environment at the time of smoking onset that might have influenced the decision to start. Moreover, never-smokers can be distinguished from ex-smokers. I estimate discrete, but also continuous time hazard models. Results indicate that young higher educated individuals are less likely to start, whereas the hazard of starting among older individuals is not affected by education. Furthermore, parental smoking during the whole childhood significantly increases the probability to start. Almost no significant effects are found regarding parental education, labor market status and living in a large city. Price effects could not be identified, because in Germany prices did not vary during the last decades up to 2002.

JEL Classification: I12

Keywords: GSOEP, youths, discrete time hazard model, log-logistic duration analysis

August 2007

---

* RWI Essen, Germany. – The author thanks Boris Augurzky, Sandra Schaffner, Christoph M. Schmidt, Marcus Tamm and Harald Tauchmann for very valuable comments. – All correspondence to Silja Göhlmann, RWI Essen, Hohenzollernstr. 1-3, 45128 Essen, Germany, Fax: +49 201 8149-200, Email: silja.goehlmann@rwi-essen.de.

1 Introduction

In Germany, awareness of the risks of smoking has been growing and overall tobacco consumption decreasing since the 1980's (Schulze and Lampert, 2006). Nevertheless, the development of smoking participation rates among teenagers gives rise to concern. In the German Mikrozensus 1992, 21% of young males between age 15-19 reported to be smokers. In 2003 this number increased to 27%. The corresponding proportions of smokers among young females between age 15-19 increased from 14% to 23% during the same period. Furthermore, the average starting age decreased from 17.6 for males born 1950-1954 to 17.1 for males born 1975-1979 (Figure 1). This decrease was steeper among women. Females born 1950-1954 started smoking on average at the age of 18.8, females born 1975-1979 at the age of 16.8.

Interestingly, the average starting age among young women born 1975-1979 does not significantly differ from that of men of the same birth years, whereas the average starting age among women born 1950-1954 is significantly higher than that of males of the same birth years. Moreover, whereas the starting age decreased almost steadily among women, among men this trend was reversed for the cohort of men born between 1965 and 1969 and between 1975 and 1979 compared to the respective preceding cohort. Since smoking at early age threatens to be particularly harmful, policy measures against smoking should target the decision to start smoking. This emphasis is all the more warranted, as smoking is addictive and, thus, it should be easier to prevent individuals from starting than to make smokers stop smoking (Douglas and Hariharan, 1994).

This raises the question of the determinants of smoking initiation. In the empirical literature, this question usually is addressed by discrete choice models or duration models. The main disadvantages of earlier studies are that they often lack contemporary socio-demographic information at the time smoking initiation took place, particularly the studies estimating duration models (see, for example, Forster and Jones, 2001; Lopez-Nicolas, 2002). Moreover, by only using information on the *current* smoking status as in studies estimating the probability to be a current smoker (see, for example, Chaloupka and Grossman, 1996; Tauras and Chaloupka, 1999; Gruber and Zinman, 2000), it is not possible to differentiate between never- and ex-smokers.

This paper combines the advantages of a rich panel data set with retrospective information on smoking initiation and contemporary information at the age when individuals started. It contributes to the existing literature in three ways. (i) For the first time, it provides results for Germany regarding the question on the determi-

nants of the decision to start smoking. The case of Germany might be particularly interesting. Previous research mainly focussed on the effect of changes in prices and regulations on the decision to start smoking. However, in Germany neither did the real cigarette price vary during the last decades up to 2002 nor any regulations. Nevertheless, smoking participation rates and the starting age changed. In consequence, other factors must have been driven the decision. (ii) By exploiting the retrospective information on smoking initiation it is possible to distinguish between never- and ex-smokers and (iii) by tracing back individuals up to the time they started smoking it is possible to account for individual characteristics at that time. That this possibility yields important insights is demonstrated by reproducing analyses of other papers which rely only on retrospective data derived from one cross-section and on current socio-demographic information.

The remainder of the paper is organized as follows. The next section provides an overview of the existing literature regarding smoking initiation. Section 3 introduces the empirical method of the paper and describes the data used for the analysis. Section 4 presents the estimation results. Section 5 concludes.

2 Literature

Smoking becomes an option for children after their tenth birthday or later, depending on their maturity and environment. From that point in their life they are under the risk of starting. In the following years, the individual will decide at each point in time on the basis of current information whether to start smoking or not. As long as they remain abstinent, the risk[1] continues to be present. Yet, as soon as the individual starts, the event of "failure" took place, thus creating a situation conducive to duration analysis. At least for longer durations of abstinence from smoking there seems to be strong negative duration dependence: the longer an individual has decided not to start smoking, the less likely it becomes that she will take up the habit. Apparently, from a certain age onwards, the hazard of starting even tends to go to zero. In conclusion, the decision to start smoking is not made at one point in time, but over and over again during a particular period of the life cycle. This period mainly seems to include the time between the age of around 10 until the beginning of the twenties.

In consequence, in order to analyze the determinants of smoking, one might seek for data that comprise information on smoking onset and as many variables as pos-

[1]The hazard of starting is defined as the probability of starting during a short period of time conditional on not having started smoking before.

sible regarding the environment of an individual, observed in the years during this particular part of the life cycle. Unfortunately, such data are usually not available. Often there is just a cross-section available, comprising information on smoking prevalence among young individuals and current socio-demographic characteristics. Thus, such data were often used to estimate discrete choice models, where the dependent variable usually is a binary indicator for smoking participation (see, for example, Chaloupka and Grossman, 1996; Chaloupka and Wechsler, 1997; Tauras and Chaloupka, 1999; Chaloupka and Pacula, 1999; Gruber and Zinman, 2000; Bantle and Haisken-DeNew, 2002).[2] Often, these approaches additionally model the conditional demand for cigarettes.

Many of these studies are based on samples from the "Monitoring The Future" project in the U.S. which only include high school students (Chaloupka and Grossman, 1996; Chaloupka and Wechsler, 1997; Chaloupka and Pacula, 1999; Tauras and Chaloupka, 1999; Gruber and Zinman, 2000). The main results of these studies regarding socio-demographic characteristics[3] are: smoking participation seems to be significantly positively correlated with current personal income from employment and other sources. It is significantly negatively correlated with being religious, living in a city, and living together with both parents. Mixed results are found for gender, age, parental education and marital status, depending on the estimation method (fixed effects model or cross-section analysis). In contrast to these studies, Jones (1989) estimates a double-hurdle model based on retrospective micro data on smoking onset and socio-demographic information, but unfortunately considers the latter only at the time of the survey. He finds that smoking onset is significantly negatively correlated with income (at household and individual level) and education, and that it follows an inverted U-shaped age profile.

For Germany, Bantle and Haisken-DeNew (2002) analyze smoking participation by estimating a logistic regression model and focus on the role of parental smoking behavior on their children's tobacco consumption. They rely on the *German Socio-economic Panel* but in contrast to the present study only on the 1999 wave. Including 16 to 19 year old youths, still living at their parents' home, the authors conclude that parental smoking behavior is significantly positively correlated with children's smoking incidence. Moreover, the probability to smoke is significantly higher for less educated, working youths or youths in apprenticeship compared to youths not working at all. It is also higher for youths with healthier parents. Less likely to

[2] Tauras and Chaloupka (1999) indeed use a panel of individuals but estimate a linear probability model on smoking participation.

[3] Chaloupka and Wechsler (1997) and Chaloupka and Pacula (1999) do not present the results regarding socio-demographic variables.

smoke are teenagers carrying social responsibility, being religious or optimistic. No significant correlation is stated for gender, parental education, income, life satisfaction and being in a relationship. Obviously, variables indicating social activities and attitudes are not exogenous. Hence, controlling for such variables might bias the results.

One disadvantage of all studies relying on current socio-economic characteristics is that a causal interpretation of these results is precluded: correlations of smoking prevalence with socio-demographic characteristics observed after someone started to smoke do not necessarily reflect the reason why someone started and stayed smoking. Furthermore, such studies do not differentiate between never- and ex-smokers. Douglas and Hariharan (1994) tellingly write: "Since the decision to continue a smoking habit involves an important consideration (current addiction) that is not present in the decision to initiate the habit, the participation elasticities that the earlier authors estimate are not equivalent to the elasticity of starting smoking".

Several authors make further use of the fact that some surveys include retrospective questions on the starting age. In detail, since these studies typically aim at analyzing the effect of cigarette prices on smoking onset, they extend these data to a panel structure by generating observations for each individual for all years up to the year when the individual started smoking. Then cigarette prices and/or anti-smoking regulations for each year were mapped onto the data set.[4] By doing so, the information is available, which cigarette price and/or regulations the individual faced when she or he started smoking. However, these studies still have to rely on the cross-section information on socio-demographic characteristics at the time of the survey. Hence, these studies only control for socio-demographic variables that are assumed to be exogenously determined before an individual decided to start smoking and do not change afterwards (Forster and Jones, 2001). Nonetheless, such a data structure allows to estimate the hazard of starting smoking conditional on not having started yet.

To estimate this hazard duration models are utilized, where duration is defined as the time until an individual starts smoking. One assumption implicit in these duration models is that eventually every individual, even if having never smoked at the time of the survey, will fail (i.e. start smoking), if life lasts long enough. However, tests show that this assumption does not hold. The model that accounts for this fact is the split population duration model as proposed by Schmidt and Witte (1989), a modified version of the standard duration model. The idea is to allow some individuals not

[4]Douglas and Hariharan (1994) do not create this kind of panel set but impute the cigarette price when individuals were 18 years old as well as the change of the cigarette price from age 15 to 18.

to fail at all and, hence, the model consists of two parts: (a) a discrete choice model to estimate the probability of being a potential smoker[5] and (b) a duration model conditional on being a potential smoker. This split population duration approach has been commonly used in the literature (Douglas and Hariharan, 1994; Jones, 1995; Douglas, 1998; Forster and Jones, 2001; Lopez-Nicolas, 2002; Madden, 2007).[6]

In sum, the results of these studies regarding the first part of the split population model indicate that the probability to be an ever-smoker increases significantly with being male or if a parent smokes. It also increases with being divorced compared to being not divorced. Yet, it is questionable if this variables captures any exogenous characteristics that are determined before starting smoking. The probability of being a potential smoker significantly decreases with education. The effect of income and age differs between studies with a tendency of not being significant. Concerning the starting age, onset is significantly delayed for older cohorts, female and better educated individuals.[7] Douglas and Hariharan (1994) and Douglas (1998) control for a variable indicating being divorced and do not find it to be significant. Estimated effects of parental smoking on the starting age are not robust, but parental smoking tends to shorten the time until someone starts smoking.

A remarkable exception to these studies estimating duration models is the analysis of DeCicca et al. (2002). This study utilizes data from the National Education Longitudinal Study 1988 that comprise information on eighth graders who were reinterviewed in 1990 and 1992, i.e. the sample includes individuals who were observed from around the age of 13-14 up to the age of 17-18. In order to analyze the determinants of smoking initiation the authors apply two different approaches. Firstly, they limit their sample to non-smoking individuals in the first wave and estimate ordered probit models where the dependent variable takes larger values the more cigarettes per day the individuals smoke four years later. Secondly, the authors use a sample that includes all eights graders (if smoking or not), among the tenth graders those who were still non-smokers in the eighth grade and among the twelfth graders those who were still non-smokers in the tenth grade. Based on this sample the authors estimate a discrete time hazard model. This model corresponds to a probit model including dummy variables indicating the grade, whereby the hazard rate is allowed to vary with age, respectively grade. In contrast to other studies, the analysis of DeCicca et al. (2002) is actually based on information on socio-demographic characteristics at each wave. Although controlling for many background characteristics

[5]Potential smoker means that the individual started smoking already or might start smoking in the future.

[6]An exceptions is Agrawal et al. (2005) estimating cox proportional hazard models.

[7]Madden (2007) finds mixed effects of education on the starting age among women.

including information on living area (urban vs. rural), family income, parental education etc., the authors report only the results regarding high school dropouts, gender, race and test score as well as cigarette tax as this is the main focus of their study.

3 Method and Data

Modeling the hazard of starting smoking requires observing the individuals at the time they are at risk of starting smoking. In contrast to most other studies, except the one of DeCicca et al. (2002), the following analysis is actually based on such a comprehensive data set. Thus, it is possible to estimate a model that is superior to approaches modeling smoking prevalence among youths in that it models the probability to start conditional on not having started smoking yet. Moreover, this model improves on duration models estimated by others by using contemporary information on socio-demographic characteristics.

Given the structure of the data set, one option to exploring the determinants of smoking is to model the dependent variable as a binary indicator that equals 1, if the individual started to smoke during this year and 0, if the individual has not started yet. Hence, estimating probit or logit models seems to be appropriate. The underlying idea of this specification is that each year an individual has to decide again, if she or he starts smoking or not. If she or he starts, then the individual is obviously not longer at risk of starting and the observations regarding this individual after this failure are removed from the sample. Obviously, this specification coincides with a discrete time hazard model, where the hazard is allowed to depend on the duration of not having started smoking yet, which in this case equals the age of the individual. This paper pays particular attention to this specification.

Nonetheless, also continuous time hazard models are estimated to compare the results with those of other authors. However, instead of estimating a split population duration model, truncated duration models are estimated on the subsample of smokers. This seems to be justified in view of the fact that almost all individuals are observed up to the time, when they were 21 years old, the age when about 90% of the individuals have started to smoke if they ever started. Within the group of continuous time hazard models, several different specifications can be estimated and tested against each other (for an introduction on duration models see Wooldridge, 2002). Hence, I estimated the exponential, Weibull, log-normal and log-logistic model. The comparison of the Akaike information criterion (Akaike, 1974) indicates that the log-logistic is the preferred distribution in accordance with the studies of Douglas and

Hariharan (1994), Jones (1995), Forster and Jones (2001), Lopez-Nicolas (2002) and Madden (2007).

For this model, the survival function (S(t)), i.e. the probability of not starting to smoke up to time t, the density function of duration T (f(t)) and the hazard function λ(t) for those individuals who are predicted to start smoking are specified as

$$S(t|\mathbf{x}_i(t)) = \frac{1}{1+\gamma t^{\alpha}}, \tag{1}$$

$$f(t|\mathbf{x}_i(t)) = \gamma\alpha t^{\alpha-1}(1+\gamma t^{\alpha})^{-2}, \tag{2}$$

$$\lambda(t|\mathbf{x}_i(t)) = \frac{f(.)}{S(.)} = \frac{\gamma\alpha t^{\alpha-1}}{1+\gamma t^{\alpha}}, \tag{3}$$

where $\gamma(t)$=exp[$\mathbf{x}_i(t)\beta$], $\mathbf{x}_i(t)$ represents a vector of time-varying observable characteristics of individual i ($i = 1, ..., N$), and α is a positive parameter (Wooldridge, 2002). This functional form captures the increase in the hazard during the early teenage years, as well as the subsequent decline as the adolescent matures.

The reason that the split population duration model has been preferred to a standard duration model is that the estimated density function of the former fits the non-parametric density function much better than the predicted density function of the latter (Douglas and Hariharan, 1994). This in turn is due to the fact that the standard duration model assumes that all never smokers, even never-smokers who are observed at an age when they are very likely not at risk of starting any longer, will eventually start smoking. Yet, restricting the sample to individuals being at an age where it is still likely to start smoking might reduce the bias that arises when estimating the standard duration model on the full sample. The advantage of this model is that it is possible to use the information on never-smokers as well. Results of this restricted standard duration model are presented in addition to the truncated duration model. Nevertheless, the predicted average starting age from this model of about 22 years still exceeds the predicted starting age of the truncated model of about 18 and the average starting age of the ever-smokers in the sample used for the continuous time duration model of about 17.

The models in this paper are estimated separately for women and men as suggested by LR-tests. Estimation results are presented for the discrete time hazard model, the truncated log-logistic duration model on the subsample of smokers and the log-logistic standard duration model on individuals not older than 21. Moreover, earlier

studies – using only information of one cross-section – are replicated by reducing the panel data set to the 2002 wave of the GSOEP and re-estimating those models based on information that was available in the year, when individuals were asked about their smoking behavior. By comparing the results the importance of contemporary covariates can be pointed out.[8]

For the empirical analysis data from the *German Socio-economic Panel* (GSOEP) were employed.[9] For more information on the GSOEP see Haisken-DeNew and Frick (2003). The GSOEP consists of 22 waves on an annual basis starting in 1984. In 2002 individuals were asked "Have you ever smoked before, i.e. have you smoked at least 100 cigarettes or other tobacco products in your life?". If the answer was 'yes', individuals were asked "How old were you when you began to smoke regularly?" and "Do you currently smoke, be it cigarettes, a pipe or cigars?". Moreover, in spring 2004 individuals were asked again if they currently smoke, be it cigarettes, a pipe or cigars. Individuals not older than 30 who did not start smoking up to 2002 but were smokers by 2004 were coded as having started to smoke in 2003.

In contrast to other studies, where only information on socio-demographic variables of one cross-section could be employed, the panel structure of the GSOEP provides the possibility of tracing back the individuals partly up to the time when individuals said to have started to smoke.[10] Although respondents were interviewed directly only after their 17th birthday, values of the included variables could be imputed (for example, the marital status is single for respondents below 17). Moreover, parents were asked about the type of school the child attended. By matching parent information to that of children, important parental characteristics like parental smoking behavior is also available.

Of course, tracing back individuals to the time they started is not possible for all individuals, even though we know their starting age. Moreover, I eliminated all observations of ever-smokers who are observed only *after* they started smoking, i.e. when they are not longer at risk. The data set used for the discrete time hazard model is further restricted to individuals being older than 12 and younger than 22 years since before and after that time almost no individual in the sample started to

[8] Clearly, comparing both results might be restricted by dealing with different sample sizes. Nonetheless, the reduced sample size ceteris paribus only leads to less significant coefficients.

[9] The data used in this paper were extracted using the Add-On package PanelWhiz v1.0 (Oct 2006) for Stata. PanelWhiz was written by Dr. John P. Haisken-DeNew (john@panelwhiz.eu). The following authors supplied PanelWhiz SOEP Plugins used to ensure longitudinal consistency, John P. Haisken-DeNew (4), Markus Hahn (1), Markus Hahn and John P. Haisken-DeNew (22), Mathias Sinning (1). Haisken-DeNew and Hahn (2006) describe PanelWhiz in detail.

[10] Individuals who stated to have started before the age of 10 were in the empirical analysis treated as having started at the age of 10.

smoke. Which observations were included in the sample is described more detailed in Figure 2. Generally, individuals could be observed only between 1984 (as indicated by the left vertical line) and 2004 (right y-axis). The right vertical line marks the year 2002, when questions about smoking onset were asked. Principally, the sample comprises only individuals who were surveyed in 2002 and answered the question on smoking onset.

To illustrate the sample structure, the following paragraphs describe five hypothetical individuals with regard to their sample affiliation. For example, individual A in Figure 2, born in 1960, would have not been included in the sample as there are no observations on this individual available, when she or he was aged between 13 and 21 years (indicated by the dark grey area) since the survey did not start before 1984. However, individual B, born in 1968, was 16 years old in 1984. Given that the individual, respectively her or his parent, was surveyed in this year, this observation would enter the sample. Further observations on individual B would also be included in the sample until the individual becomes 22 years *or* starts smoking. A similar picture emerges for individual C, born in 1978. However, any observations on this individual before 1991, when she or he became 13 years old, are not included in the sample. Observations on individual D, born in 1985, enter the sample from 1998 onwards until 2004, if available, or until the individual starts smoking. Individuals born after 1985 were not surveyed in 2002 and thus, the 1985 cohort presents the last cohort included in the sample.

The sample was further reduced by observations with missing information on at least one variable. The final numbers of person-year-observations used for the discrete time hazard analysis are 4,118 for women (representing 905 individuals) and 4,324 for men (representing 967 individuals). Out of these, 34% of the women and 38% of the men are observed as having started to smoke. Based on this sample, the percentage of individuals who started smoking at a particular age can be seen from Table 1. The highest percentage of starters is observed among the age group of 16. Out of 1,030 individuals who are observed at the age of 16 and have not started until then, 12.5% started smoking. Furthermore, Figure 3 shows the average starting age of these individuals by cohort. Compared to the average starting age of the sample not restricted to observations aged between 13 and 21 with no missing information, the starting age among the restricted sample is higher for the 1964-69 cohort, but otherwise similar. Moreover, the average starting age is throughout slightly lower for women than for men. Yet, this difference is not significant.

For the analysis the following set of explanatory variables is used: a set of year dummies; a set of age dummies (in the continuous time specification the dependence of the starting age on the duration up to failure (i.e. age) is modeled implicitly in the

regression analysis); one dummy variable indicating living in East-Germany; one being a foreigner and one living in a city with at least 100,000 inhabitants; three dummy variables regarding equivalent household income with having an income less than 700 Euro acting as reference group; six dummy variables for education (attending a higher secondary or an intermediate secondary or another type of school[11] and having a high, intermediate or basic secondary school degree with still attending a basic secondary school acting as reference group).

Explanatory factors further comprise four dummy variables for education of father and mother (i.e. parent holds a high degree or holds an intermediate degree, with parent having a basic degree acting as reference group); four dummy variables for labor market status (being in apprenticeship, being unemployed, not participating in the labor market, and having a part-time job with having a full-time job acting as reference group); two sets of five variables each, indicating that the parent was a smoker for at least one year when respondent was between 0-5, 6-9, 10-13, 14-17, and 18-21 years.[12] The variables indicating that the parent smoked when the respondent was aged between 10 and 13 is interacted with a dummy variable indicating being at least 10 years old, analogous the variables indicating that the parent smoked when the respondent was aged between 14 and 17 and between 18 and 21.

Moreover, the models include two dummy variables for marital status (i.e. one indicating living together with a partner (also married individuals) and one indicating being single but providing no information on any cohabitation, whereas the reference group comprises singles who do not live together with a partner). Because the marital status might not be exogenous, regression analyses are also carried out without these variables revealing almost no difference in the remaining results. For reasons of comparison to the results based on one cross-section as estimated by others, results are presented including the variables concerning marital status. See Table 2 for a description of the variables.

The price of cigarettes does not enter the model since prices did not vary between individuals and remained almost constant in real terms. Nevertheless, one might think of including a variable indicating the subjective price of cigarettes to an individual, where this variable would be defined as price/income. However, if income is replaced by price/income this variable only reflects a re-scaled income effect due

[11] This group also includes missing values on kind of school, when it is known that the individual still attends school.

[12] The idea is that parental smoking in the past might have formed children's future smoking behavior, maybe even differently at different points of their life. Alternatively, one might also think of controlling only for the current status of parental smoking at the time of the survey (as with all other explanatory variables).

to the lack of regional price variation. Nonetheless, also these analyses are carried out. The results reveal almost no differences in the coefficients. Results are available upon request.

Table 3 reports descriptive statistics of the observed characteristics, separately for individuals who are not observed to start smoking and on individuals who will start during the observation period. Starters are on average less educated, have less educated parents, their parents were much more often smokers during respondent's childhood and their equivalent income has been lower compared to individuals who are not observed to start smoking. Moreover, starters are more often foreigners. Finally, starters are less often full-time or part-time working individuals or are in apprenticeship, but they are more often not participating in the labor market and, surprisingly, female starters are less often unemployed than female never-smokers. Notwithstanding, these bivariate descriptive statistics do not control for several factors at a time and thus, these correlations do not need not to be confirmed by multivariate regression analyses.

4 Results

To start with, Figure 4 shows the non-parametric Kaplan-Meier estimates of the survival functions separately by gender. The survival function for women decreases sharply from the age of about 13 until the age of about 21 and then remains at a limit of about 0.5, i.e. about 50% of the female sample are estimated not to start smoking. For men a similar pattern is observed but the limit is at a lower level of about 0.4. Figure 5 shows that for both sexes the hazard of starting firstly increases until it reaches the peak at an age of around 18 years and decreases afterwards. Moreover, among young men aged around 18 the hazard of starting is almost twice the hazard of females of the same age.

Discrete time hazard model

Turning to the hazard models, the results of the discrete time hazard model are presented in Tables 4 and 5. Since some of the included variables like labor market status can not exert an influence on the probability to start smoking among young individuals, the model is estimated separately for different age groups. The results indicate that among girls from 13 to 16 years the hazard of starting increases steadily with age. Moreover, the hazard is significantly lower for high and intermediate school students, by about 3 to 6 percentage points, compared to students attending a basic

school. The hazard of starting is also lower, if the girl lives in West-Germany.

Furthermore, as indicated by tests of significance of the linear combination of coefficients regarding parental smoking, the hazard of starting increases if a parent smoked during the whole childhood of the respondent. Smoking during the first five life years of the daughter alone increases the hazard of starting by about 5 to 8 percentage points. Smoking cessation of a parent, given that the mother or father smoked for a while, when the daughter was born, does not seem to significantly affect the hazard of starting.[13] So far, the results confirm the correlations that were also indicated by the descriptive statistics. However, in contrast to the observation that girls starting to smoke have a lower income, the hazard model results indicate that the hazard of starting seems to be higher for girls with a high equivalent income of more than 1,300 Euro compared to girls with an income of less than 700 Euro. Clearly, the equivalent income need not to be the money the girls have at their disposal. Nonetheless, there might be a positive correlation between pocket money and equivalent income.

For females aged between 17 and 21, who have not started smoking up to that time, education does not seem to affect the hazard of starting any more. However, parental smoking during childhood still increases the probability to start. Moreover, among these individuals the hazard decreases with maternal education. Interestingly, among those females, who have not started up to the age of 17, the hazard of starting is increased for females with a relatively high income compared to low income individuals, whereas there does not seem to be a difference between individuals at the low and high end of the income distribution. An explanation might be that the labor market status also captures disposable income effects. Thus, results indicate that women having a full-time job face a higher hazard of starting compared to women being in vocational training. Results further indicate that if a women has not started up to the age of 17 to 18, the hazard of starting decreases from that age onwards. Overall, the hazard of starting does not seem to be affected by being a foreigner or living in a large city.

Among young men, the results show a less strong correlation with education compared to young females. High school students are indeed less likely to start, but the hazard is only 3 percentage points lower compared to basic school students. Among older males (as with older females) education does not seem to affect the hazard of starting anymore. A similar picture emerges for parental smoking. This indeed

[13]The reason that the coefficients on the parental smoking variables during the period from age 6 to 21 have to be interpreted as the additional effect of continued smoking behavior is that parents mostly started smoking before the child was born. Thus, by far the main part of parents who smoked when the child was older than 5 have also smoked, when the child was younger.

increases the hazard of starting among young males, particularly maternal smoking during puberty. However, among older males, who have not started smoking yet, the hazard seems to be affected only by paternal smoking during childhood. Results further indicate, that being a foreigner is associated with a significantly lower probability to start among young men, but with a significantly higher probability among older males. The hazard among older males also increases with living in the East and with income. Again, this contradicts the observation that male starters typically have a lower income. Interestingly, income does not seem to affect the hazard of starting among young men. Labor market status and living in a large city is not found to affect the hazard of starting. Finally, the hazard of starting seems to increase up to the age of 19 and decrease afterwards. Yet, surprisingly the hazard of starting at the age of 20 is not significantly different from that at the age of 17, whereas the hazard among males aged 19 and also that of males aged 21 is significantly higher. There is no obvious explanation for this temporary drop in the hazard.

Continuous time hazard models

The results of the truncated continuous time duration model are presented in Table 6. The main difference between this sample and the sample of the discrete time hazard model is that it only includes individuals who are observed to start smoking during the observation period. However, this time the sample is not restricted to individuals aged between 13 and 21, but now includes individuals aged between 10 and 30.[14] Results indicate, that among starters the starting age increases with education among both genders. That among women the coefficient of holding a high school degree exceeds the coefficient of still attending a high school might be explained by the structure of the data: the sample includes only ever-smokers and among them, only observations until the individuals start smoking. Thus, observations on high school graduates which are included in the sample will exhibit a higher starting age since high school graduates are typically older than high school students.

Results further suggest that the time until starting among men is shortened if the mother smoked during the whole childhood. In contrast, paternal smoking does not seem to affect the starting age of men. It does not seem to affect the starting age among women, too. Among women, also maternal smoking during the whole childhood does not seem to significantly affect the starting age. However, if the mother does not quit smoking when the daughter is aged between 10 to 13, this

[14]Nonetheless, the number of older individuals is quite low, as the sample includes only ever-starters until they start, and there are not many starters who start that late.

decreases the starting age, whereas continued maternal smoking up to the age of 14 to 17 seems to delay smoking onset for women who have not started before that time.

The labor market status is found to affect the starting age of male smokers. It indeed does not seem to affect the hazard of starting smoking, but among smokers full-time working males start later compared to males being in apprenticeship. Among women, the labor market indeed affect the hazard of starting, but among smokers does not affect the starting age. Furthermore, no significant effect is found concerning parental education, income and living in a large city. Finally, the starting age of married individuals or individuals in cohabitations does not significantly differ from that of singles not living together with a partner. Yet, the coefficient on the variable indicating being single but providing no information on any cohabitation is significant. This indicates that single individuals not answering the question on cohabitation exhibit a lower starting age compared to singles who state not to live together with a partner.

Whereas these regressions were carried out on the subsample of ever-smokers, Table 7 presents the results of analyses based on the full sample, yet restricted to individuals not older than 21. A comparison between these models suggest that among females the effect of education on the starting age is confirmed. However, income now turns out to be significantly correlated with the starting age: women with a higher income are not only more likely to start smoking but also start earlier. A similar picture emerges concerning maternal smoking during early childhood. This does not only increase the hazard of starting (at least among young females) but also shortens the time until starting. Results further suggest that individuals whose mothers hold an intermediate school degree compared to holding a basic degree start smoking later in life. Women in East Germany exhibit not only a higher probability to start but also tend to start earlier.

Among men, results confirm the significant effect of living in the East and being in training compared to having a full-time job. Results also confirm the significant effect of holding a basic and intermediate school degree. In addition, holding a high school degree and still attending a high school now turns out to significantly delay the starting age compared to basic school graduates and students. Furthermore, among individuals older than 18 maternal smoking cessation at this time further delays smoking onset.

Interestingly, whereas the regression results based on the subsample of ever-smokers do not reveal a robust significant effect of paternal smoking during the whole childhood on the starting age, regression results based on the full sample do. Smoking

onset of both genders is significantly shortened if a parent smoked during childhood.

The importance of contemporary factors

A further concern of this paper is to analyze the importance of contemporary socio-demographic information at the time individuals started smoking. Hence, a probit model explaining the probability of ever having started smoking is estimated based only on the 2002 cross-section, the cross-section where individuals were asked retrospectively about smoking onset. The dependent variable takes the value 1 if the individual is an ever-smoker and 0 if the individual is a never-smoker. The results are presented in Table 8. For women, the reset test indicates that the chosen specification is not appropriate. However, results are very similar if equivalent income is replaced by personal labor income and if the age dummy variables are replaced by age and age squared, whereas this specification is not rejected by the reset test. Hence, I will nevertheless refer to the presented estimates.

Results reveal that the probability to ever having started smoking increases with parental smoking during childhood. However, this result is the only one that this estimation has in common with the one based on contemporary information. Other previously significant coefficients drop to insignificance but instead other correlations are now found to be significant. Clearly, if significant coefficients were interpreted as significant effects on the probability to start smoking, conclusions drawn from these results would significantly differ from the conclusion drawn when contemporary information at the time of smoking onset was considered. This result also holds if the estimation is restricted to young individuals (these results are available upon request).

It is worth mentioning that such a probit model is used to estimate the probability of ever-smoking in the first step of the split population duration model (see among others Douglas and Hariharan, 1994). If this model led to biased estimates of the probabilities of smoking initiation –as shown above–, also the results of the duration part of the split population model might be biased.[15] Hence, if the assumption can be made that none of the censored observations will eventually fail, estimating a truncated duration model will be more appropriate.

When re-estimating the truncated log-logistic continuous time duration model based on information available from the 2002 wave, results indicate that most of the coefficients drop to insignificance (see Table 9), whereas some other coefficients now

[15]The reason is that in the split population duration model the likelihood of each observation is weighed with the estimated probabilities of ever-smoking (Douglas and Hariharan, 1994).

turn to be significant. Again, these results highlight the importance of using contemporary information.

5 Conclusion

In Germany, the development of smoking behavior among youths gives rise to worry, because smoking participation rates among youth steadily increased during the decade up to 2003. Moreover, individuals of later cohorts tend to start smoking earlier than older ones. When aiming at lower smoking participation rates special focus should lie on the decision of young individuals to start, since it might be easier to prevent individuals from starting the habit than to make smokers stop smoking. This paper contributes to the existing literature by providing an econometric analysis on smoking onset using contemporary information at the time individuals started as well as retrospective information and thereby allowing for a more causal interpretation of the results. Furthermore, the analysis is done for a country where price changes of cigarettes are unlikely to have caused the increase in smoking incidence rates.

By estimating hazard models, results indicate that among higher educated individuals not only the probability to start smoking is lower but also the time until failure tends to be increased. A similar effect can be observed for parental smoking during the whole childhood of the respondent. However, these effects on the hazard of starting tend to be weakened among older youths. Among older males, the decision to start seems to be driven by nationality, income and paternal smoking behavior. Among older females full-time working individuals have a higher probability to start than individuals being in apprenticeship. Moreover, women with less educated mothers are more likely to start. For both genders, no significant effects are found concerning living in a large city.

The results are different if models are re-estimated not using contemporaneous information at the starting age but using only a cross-section with retrospective information on smoking onset and information on socio-demographic variables only at the time of the survey. To conclude, because the main part of individuals starting smoking start before their 17th birthday and given that non-smoking campaigns can help preventing individuals from starting smoking, these campaigns should especially focus on individuals going to basic schools and individuals having smoking parents as they present the group with the highest risk of starting.

References

Agrawal, A., Madden, P. A. F., Heath, A. C., Lynskey, M. T., Bucholz, K. K. and Martin, N. G. (2005). Correlates of regular cigarette smoking in a population-based sample of Australian twins, *Addiction* **100**(11).

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**(6): 716–723.

Bantle, C. and Haisken-DeNew, J. P. (2002). Smoke signals: The intergenerational transmission of smoking behavior, *DIW Discussion Paper* **227**.

Chaloupka, F. J. and Grossman, M. (1996). Price, tobacco control policies and youth smoking, *NBER Working Paper Series* **5740**.

Chaloupka, F. J. and Pacula, R. L. (1999). Sex and race differences in young people's responsiveness to price and tobacco control policies, *Tobacco Control* **8**: 373–377.

Chaloupka, F. J. and Wechsler, H. (1997). Price, tobacco control policies and smoking among young adults, *Journal of Health Economics* **16**(3): 359–373.

DeCicca, P., Kenkel, D. and Mathios, A. (2002). Putting on the fires: will higher taxes reduce youth smoking?, *Journal of Political Economy* **110**(1): 144–169.

Douglas, S. (1998). The duration of the smoking habit, *Economic Inquiry* **XXXVI**: 49–64.

Douglas, S. and Hariharan, G. (1994). The hazard of starting smoking: estimates from a split population model, *Journal of Health Economics* **13**(2): 213–230.

Forster, M. and Jones, A. M. (2001). The role of tobacco taxes in starting and quitting smoking: duration analysis of British data, *Journal of the Royal Statistical Society: Series A* **164**(3).

Gruber, J. and Zinman, J. (2000). Youth smoking in the U.S.: evidence and implications, *NBER Working Paper Series* **7780**.

Haisken-DeNew, J. P. and Frick, J. (2003). *DTC Desktop Companion to the German Socio-Economic Panel Study (SOEP)*, DIW, Berlin.

Haisken-DeNew, J. P. and Hahn, M. (2006). Panelwhiz: A flexible modularized Stata interface for accessing large scale panel data sets, *mimeo* . www.PanelWhiz.eu.

Jones, A. M. (1989). A double-hurdle model of cigarette consumption, *Journal of Applied Econometrics* **4**: 23–39.

Jones, A. M. (1995). A microeconometric analysis of smoking in the UK Health and Lifestyle Survey, *Centre for Health Economics, University of York, Discussion Paper* **139**: 1–44.

Lopez-Nicolas, A. (2002). How important are tobacco prices in the probensity to start and quit smoking? An analysis of smoking histories from the Spanisch National Health Survey, *Health Economics* **11**: 521–535.

Madden, D. (2007). Tobacco taxes and starting and quitting smoking: does the effect differ by education?, *Applied Economics* **39**: 613–627.

Schmidt, P. and Witte, A. D. (1989). Predicting criminal recidivism using 'split population' survival time models, *Journal of Econometrics* **40**: 141–159.

Schulze, A. and Lampert, T. (2006). *Beiträge zur Gesundheitsberichterstattung des Bundes - Bundes-Gesundheitssurvey: Soziale Unterschiede im Rauchverhalten und in der Passivrauchbelastung in Deutschland*, Robert Koch-Institut.

Tauras, J. and Chaloupka, F. (1999). Price, clean indoor air laws, and cigarette smoking: evidence from longitudinal data for young adults, *NBER Working Paper Series* **6937**.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.

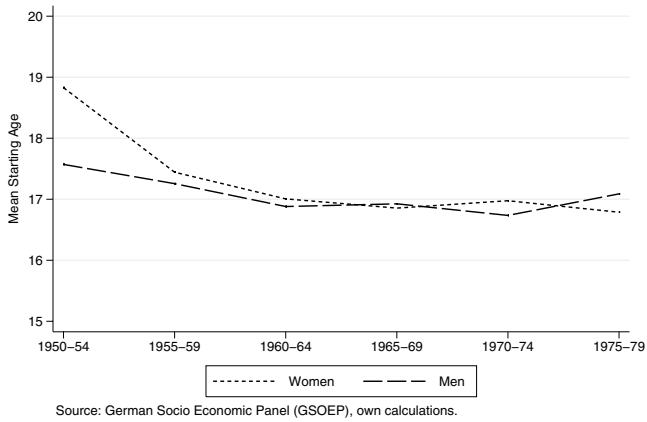

Figure 1: Development of the Starting Age by Cohort

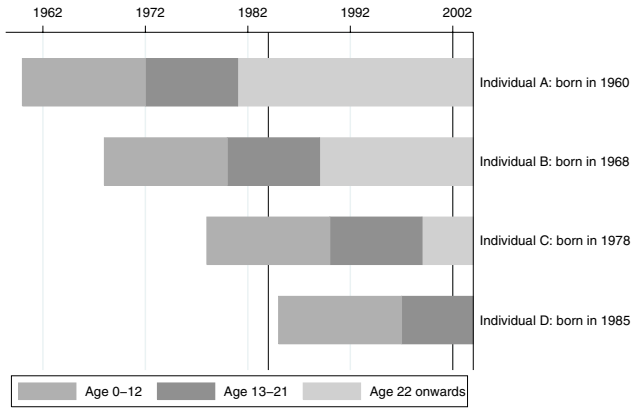

Figure 2: Structure of the Data Set

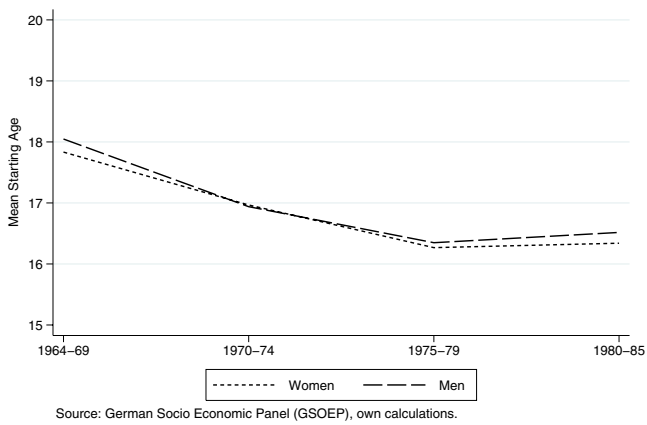

Figure 3: Development of the Starting Age Within the Sample by Cohort

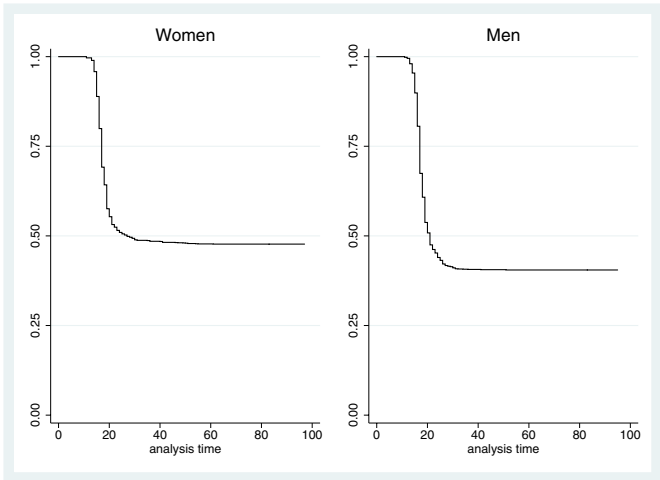

Figure 4: Kaplan-Meier Survival functions

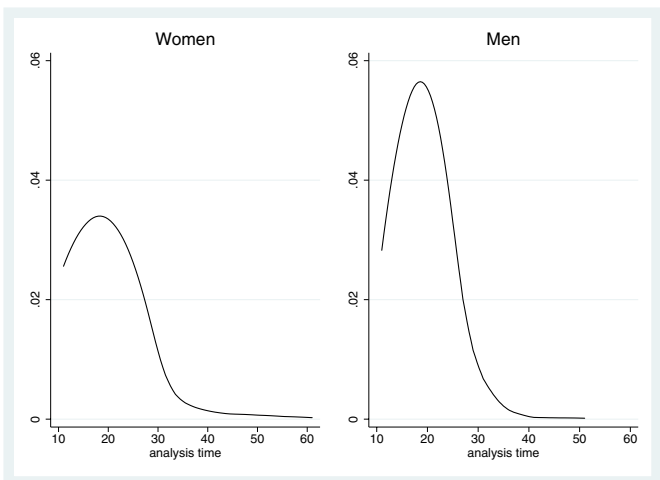

Figure 5: Kaplan-Meier Hazard functions

Table 1: **Percentage of individuals who started to smoke by age**

Age	13	14	15	16	17	18	19	20	21
Started smoking (in %)	2.5	6.1	9.0	12.5	8.5	9.3	4.2	4.2	2.1
Did not start (in %)	98.5	93.9	91.0	87.5	91.5	90.7	95.8	95.8	97.9
Total number of individuals	866	969	1,064	1,030	1,068	1,047	906	792	700

Table 2: **Description of Variables**

Variable	Description
Age	Age of individual in years
East-German	1 if individual resides in East-Germany; 0 otherwise
Foreigner	1 if individual is a foreigner; 0 otherwise
Urban	1 if individual resides in a city with more than 100,000 inhabitants; 0 otherwise
Income 700-999 Euro	1 if household equivalent income is between 700 and 999 Euro; 0 otherwise
Income 1,000-1,299 Euro	1 if household equivalent income is between 1,000 and 1,299 Euro; 0 otherwise
Income more than 1,300 Euro	1 if household equivalent income is more than 1,300 Euro; 0 otherwise
Education	
High school degree	1 if individual has a high secondary school degree; 0 otherwise
Intermediate school degree	1 if individual has an intermediate secondary school degree; 0 otherwise
Basic school degree	1 if individual has a basic school degree; 0 otherwise
Student high school	1 if individual still attends a high school; 0 otherwise
Student intermediate school	1 if individual still attends an intermediate school; 0 otherwise
Student basic school	1 if individual still attends a basic school; 0 otherwise
Student other school	1 if individual still attends another kind of school including kind of school unknown; 0 otherwise
Job	
Full-time job	1 if individual has a full-time job including civil-/military service; 0 otherwise
Part-time job	1 if individual has a part-time job; 0 otherwise
In vocational training	1 if individual is in vocational training (not student); 0 otherwise
Unemployed	1 if individual is unemployed and looking for a job; 0 otherwise
Parents	
Mother high school degree	1 if mother has a high school degree: 0 otherwise
Mother intermediate school degree	1 if mother has an intermediate secondary school degree: 0 otherwise
Mother basic school degree	1 if mother has a basic school degree: 0 otherwise
Father high school degree	1 if father has a high school degree: 0 otherwise
Father intermediate school degree	1 if father has an intermediate secondary school degree: 0 otherwise
Father basic school degree	1 if father has a basic school degree: 0 otherwise
Marital Status	
Living together with partner	1 if individual is married or single, but lives together with her partner; 0 otherwise
Not living together with partner	1 if individual is single and does not live together with a partner; 0 otherwise
Single, but living together unknown	1 if individual is single, but no information if individual lives together with her partner; 0 otherwise
Parental Smoking	
Mother smoked at age X to Y	1 if individual's mother smoked when the individual was between X and Y years old (need not to be during the whole period) and if individual was at least X years old; 0 otherwise
Father smoked at age X to Y	1 if individual's father smoked when the individual was between X and Y years old (need not to be during the whole period) and if individual was at least X years old; 0 otherwise

Table 3: **Descriptive Statistics: Discrete Time Hazard Model**

	Women				**Men**			
	Never		**Starter**		**Never**		**Starter**	
	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Age	17.228	2.400	15.566	2.070	17.358	2.430	15.707	2.141
East German	0.253	0.435	0.324	0.468	0.273	0.446	0.270	0.444
Foreigner	0.140	0.348	0.201	0.401	0.123	0.328	0.252	0.434
Urban	0.275	0.447	0.269	0.444	0.250	0.433	0.282	0.450
Income 0-699 Euro	0.125	0.331	0.139	0.346	0.118	0.323	0.162	0.369
Income 700-999 Euro	0.205	0.404	0.268	0.443	0.228	0.420	0.285	0.451
Income 1,000-1,299 Euro	0.262	0.440	0.316	0.465	0.256	0.437	0.283	0.451
Income more than 1,300 Euro	0.408	0.492	0.278	0.448	0.397	0.489	0.270	0.444
High school degree	0.090	0.287	0.025	0.156	0.082	0.274	0.019	0.136
Intermediate school degree	0.143	0.351	0.067	0.251	0.137	0.343	0.070	0.256
Basic school degree	0.069	0.254	0.048	0.214	0.126	0.331	0.089	0.285
Student high school	0.434	0.496	0.366	0.482	0.343	0.475	0.239	0.427
Student intermediate school	0.151	0.358	0.206	0.404	0.143	0.350	0.239	0.426
Student basic school	0.080	0.272	0.214	0.410	0.115	0.319	0.272	0.445
Student other school	0.032	0.176	0.074	0.262	0.056	0.229	0.072	0.259
Full-time job	0.051	0.220	0.019	0.138	0.110	0.313	0.034	0.182
Part-time job	0.031	0.176	0.018	0.131	0.032	0.175	0.009	0.095
In training	0.137	0.343	0.068	0.252	0.173	0.378	0.102	0.303
Not employed	0.768	0.422	0.888	0.316	0.675	0.468	0.843	0.364
Unemployed	0.013	0.111	0.007	0.086	0.010	0.100	0.011	0.106
Mother high school degree	0.157	0.364	0.089	0.284	0.126	0.331	0.092	0.290
Mother intermediate school degree	0.379	0.485	0.401	0.490	0.399	0.490	0.363	0.481
Mother basic school degree	0.464	0.499	0.510	0.500	0.476	0.500	0.545	0.498
Father high school degree	0.238	0.426	0.186	0.390	0.212	0.408	0.124	0.330
Father intermediate school degree	0.281	0.450	0.268	0.443	0.281	0.450	0.319	0.466
Father basic school degree	0.482	0.500	0.546	0.498	0.508	0.500	0.557	0.497
Living together with partner	0.383	0.486	0.176	0.381	0.462	0.499	0.219	0.414
Not living together with partner	0.179	0.383	0.081	0.273	0.096	0.295	0.030	0.171
Single, but living together unknown	0.438	0.496	0.743	0.437	0.443	0.497	0.751	0.433
Mother smoked at age 0 to 5	0.267	0.443	0.488	0.500	0.296	0.457	0.388	0.488
Mother smoked at age 6 to 9	0.255	0.436	0.477	0.500	0.275	0.447	0.382	0.486
Mother smoked at age 10 to 13	0.248	0.432	0.470	0.499	0.259	0.438	0.372	0.483
Mother smoked at age 14 to 17	0.215	0.411	0.375	0.484	0.230	0.421	0.304	0.460
Mother smoked at age 18 to 21	0.102	0.302	0.073	0.260	0.116	0.320	0.053	0.224
Father smoked at age 0 to 5	0.517	0.500	0.712	0.453	0.503	0.500	0.642	0.480
Father smoked at age 6 to 9	0.482	0.500	0.685	0.465	0.437	0.496	0.578	0.494
Father smoked at age 10 to 13	0.434	0.496	0.667	0.472	0.400	0.490	0.538	0.499
Father smoked at age 14 to 17	0.368	0.482	0.506	0.500	0.345	0.475	0.435	0.496
Father smoked at age 18 to 21	0.175	0.380	0.091	0.288	0.177	0.382	0.083	0.275
Number of observations	3,034		1,084		3,003		1,321	

Table 4: **Discrete Time Hazard Estimates**

	Women			
	13-16 years old		**17-21 years old**	
	Marg. effect	Std. error	Marg. effect	Std. error
Age 14	0.078 ∗ ∗	0.037	.−	
Age 15	0.113∗∗∗	0.040	.−	
Age 16	0.153∗∗∗	0.046	.−	
Age 18	.−	.−	0.002	0.013
Age 19	.−	.−	−0.023∗∗∗	0.009
Age 20	.−	.−	−0.036∗∗∗	0.008
Age 21	.−	.−	−0.043∗∗∗	0.006
East German	0.030 ∗ ∗	0.017	0.029∗∗∗	0.012
Foreigner	−0.009	0.012	−0.015	0.010
Urban	−0.008	0.010	−0.002	0.008
Income 700-999 Euro	0.012	0.017	0.024	0.019
Income 1,000-1,299 Euro	0.018	0.018	0.035 ∗ ∗	0.020
Income more than 1,300 Euro	0.050 ∗ ∗	0.023	0.017	0.015
High school degree	.−	.−	0.006	0.030
Intermediate school degree	.−	.−	−0.005	0.023
Basic school degree	.−	.−	0.014	0.029
Student high school	−0.062∗∗∗	0.014	−0.023	0.021
Student intermediate school	−0.031∗∗∗	0.010	−0.028∗	0.009
Student other school	−0.017	0.015	0.048	0.053
Part-time job	.−	.−	0.009	0.023
In training	.−	.−	−0.025 ∗ ∗	0.010
Not employed	.−	.−	−0.024	0.019
Unemployed	.−	.−	−0.021	0.015
Mother high school degree	−0.009	0.016	−0.022 ∗ ∗	0.008
Mother intermediate school degree	−0.017	0.012	−0.017∗	0.009
Father high school degree	0.025	0.019	0.007	0.013
Father intermediate school degree	−0.001	0.013	−0.003	0.009
Living together with partner	.−	.−	0.009	0.010
Single, but living together unknown	.−	.−	0.024	0.025
Mother smoked at age 0 to 5	0.081 ∗ ∗	0.041	0.027	0.030
Mother smoked at age 6 to 9	−0.092∗	0.049	−0.013	0.045
Mother smoked at age 10 to 13	0.134∗	0.101	0.027	0.068
Mother smoked at age 14 to 17	−0.017	0.020	−0.005	0.024
Mother smoked at age 18 to 21	.−	.−	0.001	0.015
Father smoked at age 0 to 5	0.048 ∗ ∗	0.023	−0.007	0.019
Father smoked at age 6 to 9	−0.048	0.043	−0.017	0.030
Father smoked at age 10 to 13	0.023	0.033	0.051	0.035
Father smoked at age 14 to 17	0.020	0.022	0.006	0.021
Father smoked at age 18 to 21	.−	.−	0.005	0.015
Number of observations	1,931		2,187	
Log Pseudolikelihood	-445.959		-401.1077	
Wald-Statistic (χ^2)	132.535		140.170	
Reset test	0.814		0.136	
Joint sign. of mother smoked 0-13	0.574∗∗∗	0.221	.−	.−
Joint sign. of father smoked 0-13	0.293	0.202	.−	.−
Joint sign. of mother smoked 0-17	0.387∗∗∗	0.108	0.376 ∗ ∗	0.168
Joint sign. of father smoked 0-17	0.485∗∗∗	0.118	0.384 ∗ ∗	0.164
Joint sign. of mother smoked 0-21	.−	.−	0.394∗∗∗	0.125
Joint sign. of father smoked 0-21	.−	.−	0.450∗∗∗	0.131

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%. Standard errors are adjusted to take repeated observations into account. Reference group for the 13-16 years old is an individual with an income less than 700 Euro and still going to a basic school with parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. Reference group for the 17-21 years old is a single individual not living together with a partner, still going to a basic school, having a full-time job and parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. The regressions further include year dummies.

Table 5: **Discrete Time Hazard Estimates**

	Men			
	13-16 years old		**17-21 years old**	
	Marg. effect	Std. error	Marg. effect	Std. error
Age 14	0.024	0.026	.–	.–
Age 15	0.066 ∗ ∗	0.032	.–	.–
Age 16	0.145∗∗∗	0.041	.–	.–
Age 18	.–	.–	0.048∗	0.030
Age 19	.–	.–	0.087∗∗∗	0.033
Age 20	.–	.–	0.023	0.024
Age 21	.–	.–	0.043 ∗ ∗	0.025
East German	0.011	0.015	0.032 ∗ ∗	0.015
Foreigner	−0.027 ∗ ∗	0.012	0.085∗∗∗	0.028
Urban	0.007	0.012	−0.008	0.009
Income 700-999 Euro	0.009	0.018	0.004	0.016
Income 1,000-1,299 Euro	0.009	0.019	0.012	0.017
Income more than 1,300 Euro	−0.011	0.019	0.035 ∗ ∗	0.018
High school degree	.–	.–	−0.032	0.016
Intermediate school degree	.–	.–	−0.035	0.017
Basic school degree	.–	.–	−0.021	0.020
Student high school	−0.029 ∗ ∗	0.013	−0.034	0.017
Student intermediate school	−0.012	0.012	0.005	0.024
Student other school	−0.022	0.014	−0.018	0.019
Part-time job	.–	.–	−0.009	0.022
In training	.–	.–	0.010	0.017
Not employed	.–	.–	−0.013	0.020
Unemployed	.–	.–	0.008	0.034
Mother high school degree	0.044	0.029	−0.002	0.019
Mother intermediate school degree	0.003	0.013	−0.008	0.012
Father high school degree	−0.007	0.016	−0.008	0.015
Father intermediate school degree	0.004	0.014	0.021	0.015
Living together with partner	.–	.–	−0.016	0.012
Single, but living together unknown	.–	.–	−0.006	0.018
Mother smoked at age 0 to 5	−0.005	0.043	0.023	0.033
Mother smoked at age 6 to 9	0.027	0.060	−0.018	0.032
Mother smoked at age 10 to 13	−0.033	0.033	0.011	0.047
Mother smoked at age 14 to 17	0.071 ∗ ∗	0.043	−0.006	0.036
Mother smoked at age 18 to 21	.–	.–	−0.009	0.017
Father smoked at age 0 to 5	−0.026	0.028	0.014	0.020
Father smoked at age 6 to 9	0.071∗	0.038	−0.017	0.031
Father smoked at age 10 to 13	0.010	0.033	0.025	0.040
Father smoked at age 14 to 17	−0.020	0.024	0.030	0.033
Father smoked at age 18 to 21			−0.014	0.014
Number of observations	1,998		2,326	
Log Pseudolikelihood	-492.232		-494.899	
Wald-Statistic (χ^2)	147.188		139.327	
Reset test	0.550		0.680	
Joint sign. of mother smoked 0-13	−0.138	0.237	.–	.–
Joint sign. of father smoked 0-13	0.479 ∗ ∗	0.215	.–	.–
Joint sign. of mother smoked 0-17	0.373∗∗∗	0.103	0.057	0.161
Joint sign. of father smoked 0-17	0.298∗∗∗	0.105	0.514∗∗∗	0.150
Joint sign. of mother smoked 0-21	.–	.–	−0.044	0.130
Joint sign. of father smoked 0-21	.–	.–	0.355∗∗∗	0.116

See notes Table 4.

Table 6: **Truncated Continuous Time Log-Logistic Hazard Estimates**

	Women		**Men**	
	Coefficient	Std. error	Coefficient	Std. error
East German	−0.0043	0.0136	−0.0375***	0.0145
Foreigner	0.0187	0.0150	0.0335**	0.0147
Urban	0.0111	0.0112	0.0047	0.0095
Income 700-999 Euro	−0.0097	0.0173	0.0043	0.0141
Income 1,000-1,299 Euro	−0.0073	0.0178	0.0131	0.0151
Income more than 1,300 Euro	−0.0162	0.0209	0.0080	0.0169
High school degree	0.2064***	0.0415	0.1455*	0.0836
Intermediate school degree	0.0716**	0.0356	0.1630***	0.0453
Basic school degree	0.0541*	0.0291	0.0745**	0.0303
Student high school	0.0468***	0.0155	0.0274*	0.0142
Student intermediate school	0.0220	0.0160	0.0183	0.0124
Student other school	0.0132	0.0210	0.0248	0.0206
Part-time job	0.0100	0.0620	0.0061	0.0520
In training	−0.0156	0.0496	−0.1335***	0.0444
Not employed	−0.0073	0.0526	−0.0926*	0.0525
Unemployed	0.1026	0.1118	0.0066	0.0562
Mother high school degree	−0.0024	0.0219	−0.0036	0.0205
Mother intermediate school degree	0.0151	0.0171	0.0027	0.0137
Father high school degree	−0.0158	0.0188	0.0150	0.0149
Father intermediate school degree	−0.0010	0.0150	0.0148	0.0141
Living together with partner	0.0019	0.0211	0.0444	0.0346
Single, but living together unknown	−0.0813***	0.0167	−0.0739***	0.0147
Mother smoked at age 0 to 5	−0.0690	0.0458	−0.0358	0.0507
Mother smoked at age 6 to 9	0.1116*	0.0655	0.0084	0.0607
Mother smoked at age 10 to 13	−0.1299**	0.0528	−0.0132	0.0513
Mother smoked at age 14 to 17	0.0659**	0.0276	0.0143	0.0388
Mother smoked at age 18 to 21	0.0222	0.0327	0.0458	0.0325
Father smoked at age 0 to 5	−0.0377	0.0360	0.0090	0.0236
Father smoked at age 6 to 9	0.0133	0.0473	−0.0043	0.0354
Father smoked at age 10 to 13	0.0187	0.0445	−0.0645	0.0456
Father smoked at age 14 to 17	−0.0182	0.0248	0.0459	0.0365
Father smoked at age 18 to 21	−0.0068	0.0249	−0.0399	0.0361
Constant	2.9056***	0.0638	3.0159***	0.0597
Number of observations	1,536		1,818	
Log Pseudolikelihood	255.833		308.695	
Wald-Statistic (χ^2)	525.434		565.981	
Joint sign. of mother smoked 0-17	−0.0215*	0.0112	−0.0264**	0.0107
Joint sign. of father smoked 0-17	−0.0239*	0.0135	−0.0139	0.0110

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%. Standard errors are adjusted to take repeated observations into account. Reference group is a single individual not living together with a partner, still going to a basic school, having a full-time job and parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. The regressions further include year dummies.

Table 7: **Continuous Time Log-logistic Hazard Estimates Based on Ever- and Never-Smokers Not Older Than 21**

	Women		**Men**	
	Coefficient	Std. error	Coefficient	Std. error
East German	−0.0575***	0.0183	−0.0337**	0.0149
Foreigner	0.0383*	0.0212	−0.0142	0.0168
Urban	0.0163	0.0157	0.0040	0.0121
Income 700-999 Euro	−0.0430*	0.0233	−0.0015	0.0188
Income 1,000-1,299 Euro	−0.0597**	0.0235	−0.0031	0.0186
Income more than 1,300 Euro	−0.0618**	0.0251	−0.0142	0.0211
High school degree	0.2426***	0.0419	0.1991***	0.0373
Intermediate school degree	0.1158***	0.0405	0.1359***	0.0407
Basic school degree	0.0600	0.0370	0.0912**	0.0400
Student high school	0.1099***	0.0221	0.0793***	0.0184
Student intermediate school	0.0598***	0.0215	0.0156	0.0166
Student other school	−0.0001	0.0310	0.0408*	0.0235
Part-time job	−0.0647	0.0596	−0.0244	0.0513
In training	−0.0485	0.0413	−0.1027***	0.0268
Not employed	−0.0615	0.0449	−0.0714*	0.0413
Unemployed	0.0310	0.0798	−0.0343	0.0560
Mother high school degree	0.0395	0.0252	−0.0190	0.0234
Mother intermediate school degree	0.0434**	0.0192	0.0004	0.0149
Father high school degree	−0.0368*	0.0223	0.0116	0.0191
Father intermediate school degree	0.0054	0.0177	−0.0201	0.0150
Living together with partner	0.0003	0.0226	0.0478*	0.0281
Single, but living together unknown	−0.1279***	0.0207	−0.1068***	0.0153
Mother smoked at age 0 to 5	−0.1247***	0.0395	−0.0072	0.0420
Mother smoked at age 6 to 9	0.1568*	0.0898	−0.0186	0.0535
Mother smoked at age 10 to 13	−0.1464*	0.0841	−0.0129	0.0464
Mother smoked at age 14 to 17	0.0220	0.0325	−0.0218	0.0357
Mother smoked at age 18 to 21	0.0152	0.0300	0.0814***	0.0260
Father smoked at age 0 to 5	−0.0411	0.0370	−0.0028	0.0289
Father smoked at age 6 to 9	0.0671	0.0559	−0.0167	0.0398
Father smoked at age 10 to 13	−0.0811	0.0500	−0.0509	0.0385
Father smoked at age 14 to 17	−0.0543*	0.0292	0.0050	0.0301
Father smoked at age 18 to 21	0.0415	0.0261	0.0325	0.0213
Constant	3.1309***	0.0681	3.1685***	0.0562
Number of observations	5,029		5,166	
Log Pseudolikelihood	-123.505		-114.522	
Wald-Statistic (χ^2)	499.616		646.919	
Joint sign. of mother smoked 0-17	−0.0923***	0.0182	−0.0605***	0.0142
Joint sign. of father smoked 0-17	−0.1094***	0.0198	−0.0655***	0.0147

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%. Standard errors are adjusted to take repeated observations into account. Reference group is a single individual not living together with a partner, still going to a basic school, having a full-time job and parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. The regressions further include year dummies.

Table 8: **Discrete Choice Estimates (ONLY 2002-Wave)**

	Women		**Men**	
	Marg. effect	Std. error	Marg. effect	Std. error
East German	0.098 ∗ ∗	0.044	0.061	0.039
Foreigner	−0.044	0.065	0.135 ∗ ∗	0.054
Urban	0.006	0.038	0.033	0.036
Income 700-999 Euro	−0.021	0.080	−0.117	0.084
Income 1,000-1,299 Euro	0.023	0.079	−0.075	0.079
Income more than 1,300 Euro	0.051	0.073	−0.047	0.075
High school degree	−0.276∗	0.144	−0.104	0.142
Intermediate school degree	−0.096	0.161	0.099	0.142
Basic school degree	0.101	0.168	0.202	0.135
Student high school	−0.172	0.138	−0.187	0.125
Student intermediate school	0.039	0.177	−0.072	0.150
Student other school	0.140	0.198	−0.196	0.148
Part-time job	0.104	0.066	0.151 ∗ ∗	0.068
In training	−0.034	0.066	0.058	0.057
Not employed	−0.035	0.055	0.033	0.057
Unemployed	0.179	0.112	0.085	0.074
Mother high school degree	0.039	0.062	0.007	0.061
Mother intermediate school degree	0.031	0.046	−0.041	0.042
Father high school degree	0.006	0.055	−0.009	0.052
Father intermediate school degree	0.010	0.049	0.047	0.043
Single	−0.018	0.071	−0.096∗	0.057
Single*living together with partner	0.128∗∗∗	0.039	0.039	0.036
Single*living together unknown	−0.184	0.131	−0.082	0.157
Separated, divorced, widowed	0.359∗∗∗	0.096	0.293 ∗ ∗	0.110
Mother smoked at age 0 to 5	0.091	0.099	0.063	0.088
Mother smoked at age 6 to 9	0.224	0.155	−0.018	0.139
Mother smoked at age 10 to 13	−0.267∗	0.145	−0.159	0.154
Mother smoked at age 14 to 17	0.171	0.136	0.383∗∗∗	0.122
Mother smoked at age 18 to 21	0.064	0.093	−0.189∗	0.094
Father smoked at age 0 to 5	0.041	0.080	0.037	0.067
Father smoked at age 6 to 9	−0.190∗	0.112	0.044	0.094
Father smoked at age 10 to 13	0.276 ∗ ∗	0.116	−0.056	0.108
Father smoked at age 14 to 17	0.077	0.108	0.123	0.103
Father smoked at age 18 to 21	−0.040	0.072	−0.010	0.070
Number of observations	997		1,250	
Log Pseudolikelihood	-561.832		-755.167	
Wald-Statistic (χ^2)	224.467		200.212	
Reset test	0.0053		0.2344	
Joint sign. of mother smoked 0-17	0.522 ∗ ∗	0.219	0.735∗∗∗	0.238
Joint sign. of father smoked 0-17	0.538∗∗∗	0.179	0.373 ∗ ∗	0.171

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%. Reference group is a single individual not living together with a partner, still going to a basic school, having a full-time job and parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. The regression further include age dummies.

Table 9: **Truncated Continuous Time log-logistic Hazard Estimates, ONLY 2002-Wave**

	Women		**Men**	
	Coefficient	Std. error	Coefficient	Std. error
East German	0.01736	0.01847	−0.00609	0.01329
Foreigner	0.04901∗	0.02678	0.08174∗∗∗	0.01887
Urban	0.01009	0.01653	0.00724	0.01280
Income 700-999 Euro	0.03541	0.03285	0.06161∗	0.03264
Income 1,000-1,299 Euro	0.02612	0.03292	0.02931	0.02802
Income more than 1,300 Euro	0.02903	0.03185	0.04975∗	0.02610
High school degree	0.08724	0.07164	0.01546	0.04156
Intermediate school degree	0.05863	0.06934	−0.00485	0.04006
Basic school degree	0.01668	0.06907	−0.03294	0.03948
Student high school	0.09284	0.06794	0.00953	0.04143
Student intermediate school	0.13392∗	0.07194	−0.03426	0.05293
Student other school	0.09529	0.06870	0.03404	0.07143
Part-time job	−0.02842	0.02584	−0.00567	0.02053
In training	−0.00534	0.02450	−0.00484	0.01699
Not employed	−0.02813	0.02384	0.01653	0.01983
Unemployed	0.00073	0.03428	−0.00610	0.03026
Mother high school degree	0.00347	0.02487	−0.00658	0.02559
Mother intermediate school degree	−0.00395	0.01925	−0.01030	0.01390
Father high school degree	−0.01190	0.02251	−0.00749	0.02114
Father intermediate school degree	0.00033	0.01782	0.02255	0.01529
Single	0.06747∗∗	0.03199	0.02319	0.02044
Single*living together with partner	−0.00343	0.01768	−0.00323	0.01231
Single*living together unknown	−0.19145∗	0.11158	0.04130	0.07595
Separated, divorced, widowed	0.06601	0.04673	−0.05123∗	0.03004
Mother smoked at age 0 to 5	−0.06953∗	0.04179	−0.02789	0.03003
Mother smoked at age 6 to 9	0.00894	0.05494	0.02229	0.04930
Mother smoked at age 10 to 13	0.00345	0.04707	−0.03109	0.05906
Mother smoked at age 14 to 17	0.00862	0.04512	0.03906	0.05323
Mother smoked at age 18 to 21	0.01956	0.04147	−0.01322	0.03457
Father smoked at age 0 to 5	−0.08807∗∗	0.03670	−0.00018	0.02379
Father smoked at age 6 to 9	0.00534	0.07367	−0.00881	0.02843
Father smoked at age 10 to 13	0.08059	0.07141	−0.04285	0.04100
Father smoked at age 14 to 17	−0.02531	0.03460	0.04394	0.04046
Father smoked at age 18 to 21	0.00707	0.02346	−0.01479	0.02295
Constant	2.62702∗∗∗	0.07740	2.65695∗∗∗	0.05009
Number of observations	421		632	
Log Pseudolikelihood	254.003		387.666	
Joint sign. of mother smoked 0-17	−0.049	0.040	0.002	0.033
Joint sign. of father smoked 0-17	−0.028	0.025	−0.008	0.024

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%. Standard errors are adjusted to take repeated observations into account. Reference group is a married individual, still going to a basic school, having a full-time job and parents both having a basic school degree and whose parents did not smoke during childhood of the respondent. The regressions further include age dummies.