# Smoking on the shoulders of giants
# - An application of the two-part model to Swedish longitudinal data

Gustav Kjellsson

Master's Thesis, Lund University, Department of Economics
Supervisors: Carl-Hampus Lyttkens & Ulf-G Gerdtham

Sept 2008

# Abstract

As the title implies this study builds upon earlier research in the area to obtain an increased understanding of the determinants of smoking. Starting from the household production model (Becker 1965; Grossman 1972), the decisions of smoking participation and intensity of cigarette consumption are estimated in two sets of Two-part Models (2PM). Due to the characteristics of the data – the first three "health waves" of the Swedish Survey of Living Conditions (ULF) – both a dynamic and a static specification of the models are employed.

The study finds that irrespective of how addiction is measured, past behavior strongly affects the smoking decisions. Among the other findings, the effect of aging is shown to have a pattern of an inverse u-curve. Moreover, full income does not affect the decisions, whereas experiencing financial stress has a positive effect throughout the models. The negative effects of higher education and having several children living in the household go primarily through the participation decision.

Keywords: *Smoking; Two Part Model; Survey of Living Condition (ULF); Longitudinal data*

# Contents

# 1 Introduction & Background

## 1.1 Introduction

Substance use in general, and smoking in particular, is regarded as one of the main risk factors of health. A number of serious illnesses, including lung cancer and cardiovascular diseases, are caused and aggravated by smoking. Reducing the smoking prevalence and cigarette consumption is one of the main goals of the Swedish public health policy (SCB 2007; Folkhälsoinstitutet 2007). To be able to affect the smoking decisions by the implemented policies, it is of the highest importance to understand the determinants behind the smoking behavior of individuals.

### *1.1.1 Aims and purpose*

The primary aim of this thesis is to develop a model, and evaluate the relative importance of the determinants of smoking behavior in a Swedish context. In a larger perspective this thesis is a part of a research project examining the role of an aging population on health inequalities, and is a first step in a further analysis of inequalities in health and health investments using decomposition of Concentration Indices. By following a cohort in three consecutive waves during 1980-1997, the consequences of the aging population will be evaluated.

To decompose Concentration Indices of the risk factors of health – such as smoking – an underlying regression is needed (see O'Donnell et al 2008). The further aims of this study are therefore to develop a model and produce robust estimates that can be used in the future decomposition exercise. In some areas, this will constrain the study. Moreover, the study intends to build upon earlier economic work on smoking behavior; both theoretically and empirically. Therefore it is not a disadvantage that economists already have attempted to model the smoking decisions of individuals.

### *1.1.2 Delimitations of the study*

To be able to derive benefits and costs of smoking, the household production model is used as a base for a theoretical framework, where smoking is treated as a derived demand (compare Grossman 1972; Becker 1965). Econometrically, a 2 Part Model is applied to estimate participation and conditional consumption separately. The lag of the dependent variable and the number of years smoked are employed as measures of addiction in separate models. A pooled and a panel version of the two models are performed on the unbalanced panel, which includes individuals aged 20-68 in 1980/81. The exclusion of younger individuals makes income a better indicator of the living conditions, which is preferable in the decomposition exercise.

In contrast to the rational addiction literature, no indicator of future consumption is considered. Testing if individuals' behavior is rational or myopic is out of the scope of this paper. Due to computational difficulties interaction terms have been left out of the models, this restricts the marginal effects to be constant over time and between different groups of individuals(compare Ai & Norton 2003). This will be considered in future work. The same applies for distinguishing between quitting and starting in the participation decision.

### *1.1.3 Contributions to earlier literature*

Although the main purpose of the study has aims beyond this thesis, the study itself is of highest interest and contributes to the earlier literature alone. This study combines the use of micro level panel data and a solution to handle the problem of mass zero outcomes. There is no earlier study on Swedish data taking such an ambitious approach to model the dynamics of the smoking behavior of individuals. Lindström and colleagues (e.g. Lindström et al 2000; 2003a; 2003b) have used logistic regressions in a number of studies in a cross sectional setting. In several publications Bask and his co-authors (e.g. Bask & Melkersson 2003; 2004) use the advantage of panel data, but on an aggregated level. Lundborg & Lindgren (2004) use a two-part model to take account for the mass zero outcome on a cross-sectional dataset including only adolescents.

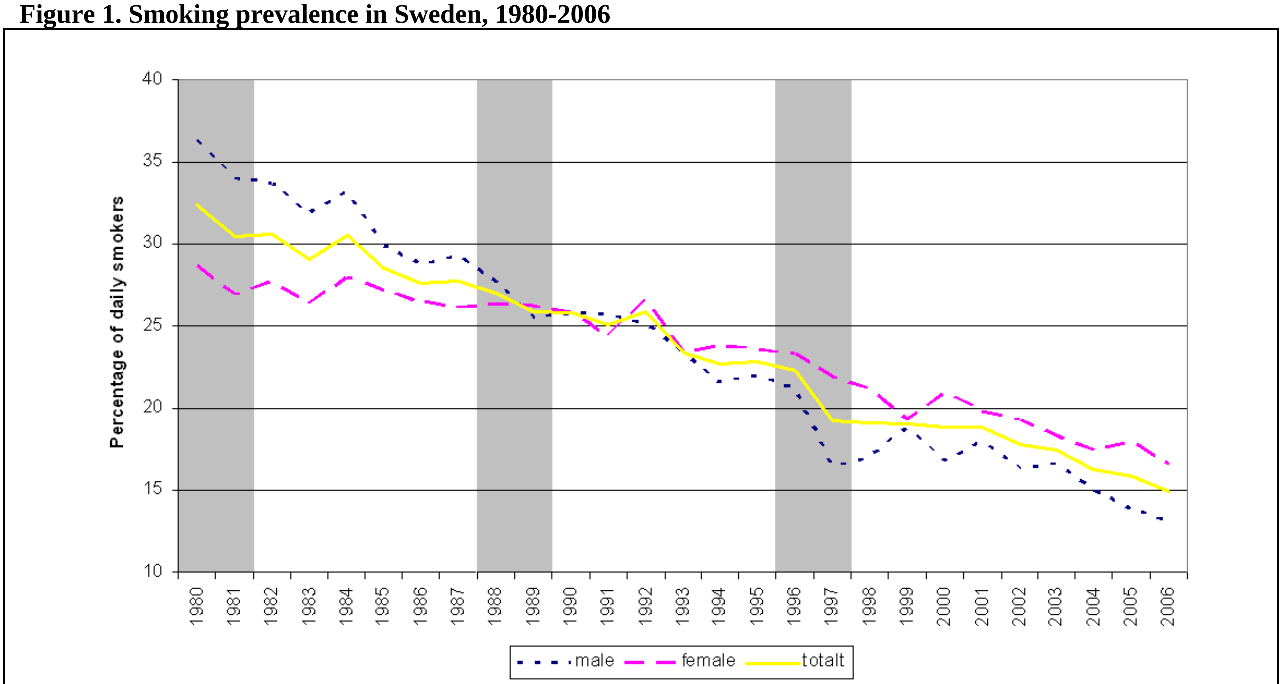
In a global context, the study contributes with new estimates of the participation probability, the conditional quantity and the total marginal effects of smoking in a Two-part Model. There are a limited amount of earlier studies combining the hurdle approach and panel data. Moreover, to the best knowledge of the author this is also the first study in this setting using a Mundlak type specification to parameterize the individual effects (compare Mundlak 1978, Jones et al 2007).

## 1.2 Smoking in a Swedish context

During a period of 30 years after the Second World War cigarette consumption increased rapidly in Sweden, from 500 cigarettes per person in 1946 to the peak of 2000 in the period of 1976-1980. Since then the amount of cigarettes smoked has decreased to 1100 per person in 2000. Smoking prevalence shows a similar pattern, in 1946 half of the men and a tenth of the women smoked regularly. During the peak of cigarette consumption, 1976-1980, 42% of all men and 34% of the women were regular smokers (Nystedt 2006).

The decline of smoking prevalence over the past 25 years is illustrated in Figure 1. The total percent of daily smokers has decreased from 32% in 1980 to 15% in 2006. As seen, the decline is largest among men, going from 36.3% to 13.2%, while female prevalence has decreased from 28.7% to 16.6%. This indicates a change in the smoking pattern between the sexes. In an international perspective, Sweden became the first country to achieve the World Health Organization's goal of reducing overall smoking prevalence to 20% (Nystedt 2006).

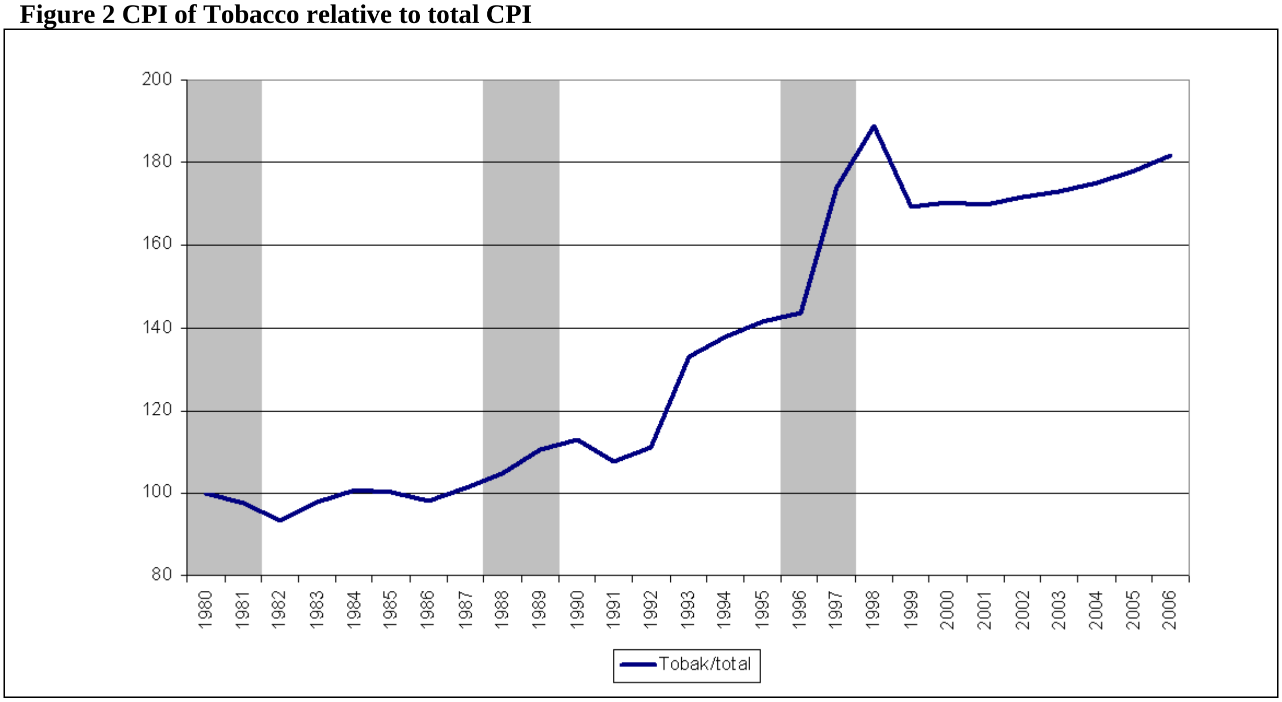
**Figure 1. Smoking prevalence in Sweden, 1980-2006**




*Source; Statistic Sweden, www.scb.se*
*Note: The data is aggregated ULF-data (the same survey as the data used in the study) directly extracted from Statistic Sweden's homepage. The grey zones indicate the years included in this study. The population is individual aged 16-80*

During the same time, the relative price of tobacco, measured as the Consumer Price Index (CPI) of Tobacco divided by total CPI, has increased significantly. In the 1990's when the overall inflation in Sweden was declining, cigarette prices was still increasing primarily due to tax raises (Hammar & Carlsson 2001). The most dramatic change is between 1996 and 1997, which constitute the third wave in the sample, a tax raise increased prices from of a package 31 SEK to 45 SEK[1] (Hammar & Carlsson 2001). If one compares the diagrams of Figure 1 and Figure 2, the participation decision seems to be responsive to prices. Especially, male prevalence appears to decline when the relative price increases.

**Figure 2 CPI of Tobacco relative to total CPI**




*Source: Statistic Sweden, www.scb.se*
*Note: 1980 is the base year in both series. The series are yearly averages of CPI. The grey zones indicate the years included in this study.*

## 1.3 Disposition

The next chapter presents an overview of earlier literature to put this study in a context. Based on earlier work in the area, a theoretical framework is developed in Chapter 3. In Chapter 4, the data and the descriptive statistics are presented, and the included variables are explained and justified. Thereafter, Chapter 5 works out the empirical modeling and explains the employed econometric

[1] The first of January 1997 cigarette prices first increased by 18%, 1 of August 1997 by additional 22%.

methods. Further, the estimated models and their results are presented and discussed in Chapter 6. Finally, in the concluding remarks future research are elaborated.

# 2 Literature review

To put this paper in a context, a review of the earlier literature is presented. The following chapter discusses the theoretical and empirical modeling of smoking as well as reviewing the main Swedish studies.

## 2.1 Theory overview

In the economic literature, there is much evidence that the price of cigarettes affects the smoking decisions, showing that smoking is a relevant issue for economics. However, from an economic point of view the monetary price is just one part of the shadow price of consuming cigarettes. During the second half of the last century, economist started to analyze substance-use in a different way than other goods, considering the addictiveness in the economic framework.

### *2.1.1 Early work*

In a series of articles, Pollak (1970; 1976; 1978) argues that the preferences of individuals are endogenous and partly a function of past consumption; earlier consumption of a good affects either the physiologically or psychologically necessity of the good. The habit formation is created step by step, since the individual is myopic and does not consider that his consumption today affects his future tastes. However, Pollak allows the individual to recognize the negative health effects of smoking. Becker & Stiegler (1977) develops a capital accumulation approach. In terms of the effect of earlier consumption, the difference from Pollak's approach is mostly semantic. For example, exposure to certain music has a positive effect on the demand for that type of music in the future; it does not matter whether past consumption affects future demand through a change in preferences or as an accumulation of capital.

The myopic models of Winston (1980), Shelling (1984) and Thaler & Shefrin (1981) are all similar in the way that an individual's decision-making process is an on-going battle between two different "persons" or "wills" of the individual; One forward-looking that maximizes life-time utility and one short-sighted. Even if the forward-looking planner is in charge for the majority of the time, the short-sighted only needs to be in control for very short periods to ruin the long term plan of the other. In Akerlof (1991), a lack of self-control explains the myopic behavior. Individuals' tendencies to always postpone the cessation decision to tomorrow is understood as a series of repeated decisions with a small cost, where each individual decision still can be close to maximization, but where the cumulative loss can be extensive.

### 2.1.2 Rational Addiction

Two decades ago, Becker & Murphy (1988) presented the idea of the rational and "happy" addict, which, analogously to the consumption of other goods, chooses his level of substance-use to maximize his own utility over time. To characterize addiction Becker & Murphy refer to three psychological concepts. *Reinforcement* means that past (present) consumption increases the marginal utility of the present (future) consumption. The second concept, *tolerance* implies that increased past consumption yields a less satisfying feeling of consuming a given level of the good. Lastly, *withdrawal* refers to the well documented negative psychological and physical reactions of quitting or reducing the level of smoking (see e.g. Doll 2000, Surgeon General 1989). In the model, the smoking decisions are based on the benefits and the shadow price of smoking, including the market price and the future costs. As a fully rational actor, individuals do not only acknowledge that smoking has a negative effect on future utility and earnings through the reduced health, but they do also consider the habit forming or addictiveness of smoking. Following Stiegler & Becker (1977), past consumption influences the current pattern through a process of "learning by doing". Becker & Murphy contend that it would therefore be naive to allow individuals to not consider the current consumption's effect on the utility of future consumption.

### 2.1.3 Critics of Rational Addiction

In reality, many addicts are regretful. Orphanides & Zervos (1995; 1998) present the idea that individuals can be rational AND regret their earlier decisions by introducing uncertainty. Individuals are fully aware of the harmfulness of smoking, but in contrast to the view of Becker & Murphy (1988), they do not wish to become addictive. However, individuals cannot completely identify the risks -the addict-proneness of themselves- and therefore do not believe that they will become addicted. Consequently, they are willing to expose themselves to the risk of becoming addicted in the future in order to consume the good in the present. Naturally, individuals who “get hooked” will be unhappy ex post. Further, the myopic behavior of smokers is in this context explained as a consequence -rather than a cause -of smoking. Time preferences is not seen as constant, instead they are internalized and affected by earlier consumption; the more an individual consume of the addictive good, the more myopic his time preferences become.

Analogously to Simon (1978), Suranovic et al (1999) assume the individual to be "boundedly" rational. In contrast to Becker & Murphy (1988), the individual does not maximize life-time utility, instead he (or she) only chooses how much to consume today. While allowing the individual to consider the future health effect (compare Pollak 1970; 1976), Suranovic et al (1999) do not allow the individuals to perform superhuman calculations or to recognize the future -presently not known- prices that are needed to make fully consistent decisions when maximizing life-time utility.

Taking a starting point in the rational addiction model, Gruber & Köszegi (2001) contend that individuals are forward-looking in their smoking decision, but simultaneously the authors include inconsistent time-preferences in their model. In similarity with Laux (2000), the authors argue for the presence of a market failure of addictive or intrapersonal externalities, when individuals fail to internalize future non-monetary costs in their present decision.

## 2.2 Empirical modeling

The earlier empirical literature of smoking is primarily dominated by regulatory issues and the question of whether the price affects the use of tobacco. The negative price effect seems to be one of the most consistent findings throughout the literature (Ross & Chaloupka 2001). For a more extensive survey see Chaloupka & Warner (2000).

### *2.2.1 Test of Rational addiction*

During the last two decades, the proportions of papers modeling the smoking decisions in general or testing whether individuals are myopic or rational have boosted. The earliest tests of the rational addiction hypothesis were conducted by Chaloupka (1991) on micro level data and by Becker et al (1994) on aggregated macro data. To test the rational model against the myopic a lead of consumption is included. To deal with the problem of endogeneity, past and future prices are suggested as instruments. Both of the early tests, finds the lead to be significant, supporting the hypothesis of a rational addict.

The tests have been followed up by several other scholars and the findings have fairly consistently supported the rational model above the myopic. However, these tests have been criticized in the literature. Critics have both shown the test to be sensitive for empirical modeling as well as found goods as milk and eggs to be addictive (see Wangen 2004; Kan 2007; Auld and Grootendorst 2004; Gruber & Köszegi 2001).

### *2.2.2 The smoking decisions and the hurdle approach*

In the empirical modeling of the smoking four main decisions appear: Initiation, cessation, participation and quantity. Depending on the data of the study and the decision of interest, a set of different approaches and techniques to analyze smoking behavior is employed throughout the literature. In a cross-sectional setting, most of the recent studies are conducted using a hurdle or two-part approach. The smoking decisions of participation and quantity are divided into two separate decisions; to be observed with a positive outcome one need to pass the participation hurdle (see Jones 1989; Garcia & Labeaga 1996; Yen & Jones 1996; Ross & Chaloupka 2001; Tauras 2005; Chen & Farrelly 2001; Raptou et al 2005; Chaloupka & Wechsler 1997; Stehr 2007).

The hurdle approach is also applied to other combinations of the decisions. Jones (1994) estimates two sequential binary models for attempts and success, given an attempt, of quitting. The results stress the importance of the influence of social interaction, earlier smoking patterns and health status on smoking behavior. With datasets including the appropriate information, duration models to estimate the propensity to start or/and quit given the characteristics of the individuals are commonly employed (e.g. Foster & Jones 2001; Lopez-Nicolas 2001; Tauras & Chaloupka 1999)

### *2.2.3 Quitting vs. starting*

There is also a debate in the literature surrounding how to model participation. Reasonably, the decisions to start and quit differ from each other. In Jones (1989), where the author develops the double hurdle model for smoking decisions, a trivariate model -in contrast to the bivariate- with two different participation equations is discussed but due to computational issues not presented. Instead sample separation techniques are used. Now the individual needs to pass two hurdles before a positive outcome is observed. One has to both be a potential smoker and actually smoke or purchasing the good in a specific time period to have a positive consumption or level of expenditures.

In Yen and Jones (1996) the hurdle model is extended to consider that addiction affects participation and consumption asymmetrically. Using the auxiliary information of past smoking pattern, the sample is restricted to current and ex- smokers. The analysis is then conducted by a Box-Cox double hurdle to nest the standard versions of the double hurdle-model.

### *2.2.4 Panel data modeling*

Even though the advantages of longitudinal data when modeling smoking behavior are frequently stressed, the amount of studies conducted is limited. Among the few publications, Labeaga (1993, 1999) solves both the endogeneity problem in the rational addiction model and the problem of mass zero outcomes using advanced panel data methods in two steps. In addition, he accounts for unobserved heterogeneity using an approach suggested by Chamberlain (1984).

Jones & Labeaga (2003) combines the longitudinal techniques from Labeaga (1993, 1999) with the sample separation used by Jones and his coauthors (Jones 1989, 1994; Yen and Jones 1996). In contrast to Labeaga (1999) who assumes one common latent variable to decide participation, this analysis distinguishes the ones that never smoked from the sample of potential smokers. In relation to this, DeCicca et al (2008) contend that in a longitudinal setting the problem can also be solved by the inclusion of interaction terms with past consumption in a dynamic participation equation. The authors model addiction in a myopic setting stressing the importance of considering the different character of the decisions of initiation and cessation. In a recent paper, Gilleskie & Strumpf (2005) attempt to distinguish between state dependency, or addiction, and unobserved individual heterogeneity in a myopic model. The findings support the importance of state dependency among young adults.

## 2.3 Swedish studies

Lindström and colleagues (Lindström et al 2000; 2003a; 2003b; Lindström & Östergren 2001; Lindström & Sundquist 2002; Lindström 2004; 2007; Janzon et al 2005; Lindström & Janzon 2007) have in several publications examined different issues of smoking using logistic regression methods on several data sets from Southern Sweden. The majority[2] of the publications focus primarily on the correlation and causality of education, social participation and social capital on the smoking decisions; cessation, participation and initiation.

However other issues are investigated as well. In a recent paper, Lindström (2007) shows that men tend to use moist snuff – snus – as a support to quit smoking, while women mostly use nicotine replacement therapy. These findings are confirmed by Foulds et al (2007) who argue that snus has significantly contributed to the decline in smoking prevalence among Swedish men.

In a similar context, Bask & Melkersson (2003) argue in a criticized article (see Ault et al 2005) moist snuff to be an inefficient tool in smoking cessation. The applied rational addiction model is extended to enclose two addictive goods as developed in Bask et al (2006). The authors apply the same model to aggregated data of legal and illegal cigarettes in Bask et al (2003) as well as cigarettes and alcohol in Bask & Melkersson (2004). In the latter article, a test of the Rational Addiction hypothesis is conducted using both separate and system of equations on time-series data of sales volumes from the second half of the last century. While the demand for alcoholic beverages appears consistent with the rational addiction model, the same hypothesis is rejected for cigarette demand.

Using the same data source as this study, Nystedt (2006) shows the protective effect of marriage on smoking. In order to study the dynamics of the marital life course closer, Nystedt performs a set of logistic regressions to examine the effect of a change in marital status on initiating and cessation. The main findings are that marriage overall protect both men and female from smoking, while marital disruption has the reverse effect.

---

[2] Lindström et al 2000; 2002; 2003a; 2003b Lindström & Östergren 2001, Lindström & Sundquist 2002, Janzon et al 2005;, Lindström & Janzon 2007

In a number of publications from his dissertation Petter Lundborg investigates drug-use among Swedish adolescents. Lundborg uses data from three surveys conducted in Trelleborg in the South of Sweden from 1999-2001. Two of these papers touch upon smoking behavior. First, Lundborg & Lindgren (2002) use a negative binomial hurdle model to analyze the connection between risk perceptions of smoking related lung-cancer and cigarette consumption. The results indicate that adolescents tend to overestimate the risk of smoking. Further, the perceived risk has a negative effect on the participation decision, but appears to have no influence on the intensity. In the second paper, Lundborg (2004) finds significant peer effects on the smoking behavior of adolescents, an effect that appears to be stronger among females.

# 3 Theoretical framework

## 3.1 The Economic Household model

Starting from standard economic theory, this paper assumes that individuals act according to their preferences, utility functions and constraints. Further, this paper advocates that the smoking decisions are based on the perceived benefits and costs. As Becker & Murphy (1988) conclude, the shadow price of smoking includes both the monetary price and the non-monetary costs, which first and foremost are health-related.

In order to derive the benefits and the shadow price of smoking, a theoretical framework needs to be developed. Using the household production model (compare Grossman 1972, Becker 1965; Rosenzweig & Schultz 1983) as a starting point, individuals enjoy utility from the consumption of different goods or commodities. Further, individuals are not only consumers of the commodities, but also producers, using inputs purchased at the market and own time in the production function. Cigarettes or smoking is therefore best treated as an input in different production functions. In this way the demand for cigarettes is a derived demand; cigarettes is not directly included in the utility argument of the model, but indirectly since it produce commodities that yield utility (compare DiCecca et al 2000).

The benefits - or reasons- of substance use can be divided into internal and external. Internal physiological (or psychological) effects of smoking as relaxation will further on be referred to as intoxication. Reasonably, there are also external social factors behind smoking patterns, cigarettes can make you perceive yourself as looking "cool" or help you to fit in to a group. Consequently, in the framework of a household production model cigarettes produce peer or social acceptance and intoxication (compare Emery et al 2001 & Powell et al 2005). Naturally, individuals also enjoy being healthy and it is therefore reasonable to directly include health in the utility function. Hence, in the economic framework of the paper, the utility argument of the model includes; Health, Intoxication, Social or Peer Acceptance and all other commodities.

To summarize, smoking affects health negatively, it can produce a coolness capital or peer acceptance, and act as an input in the production function of intoxication. Just as the demand for health vary among individuals due to their opportunity cost and certain characteristics (compare Wagstaff 1986; Muurinen 1982; Grossman 1972), so does the demand for social acceptance and intoxication.

## 3.2 Rational addiction concepts in the Household production model

When implementing the psychological terms of the rational addiction model in the context of a household production model -smoking is now treated as an input in the production function of an addictive commodity rather than an addictive good in itself- tolerance affects the supply side, reducing the productivity of intoxication. A higher level of cigarette consumption is needed to produce the same level of intoxication. Simultaneously, reinforcement -or the addictiveness of the commodity- has an effect on the demand side as past consumption increases the marginal utility of intoxication today. In terms of the rational addiction model, this implies that intoxication today is a close compliment to past intoxication. The third concept, Withdrawal -the physical cost of quitting or reducing the amount of cigarettes- shall be interpreted as an additional strengthening of the reinforcement effect on the demand for intoxication.

In similarity with Orphanides & Zervos (1995; 1998) and Suranovic et al (1999), the author of this paper do not believe in the happy addict, but in a somewhat rational agent; since individuals still consider the benefits and the costs of their decisions. If this is explained by the uncertainty of the person's addict-proneness or by the boundedly rational individual of Simon (1978), is only a matter of semantic (see Orphanides &Zervos 1995; 1998; Suranovic et al 1999). Furthermore, even though individuals acknowledge the negative future health effects of smoking, it is to costly to perform the superhuman backwards inductions needed to consider the second-order effect of smoking (compare Suranovic et al 1999). In other words - with the risk of being accused of naïveté- this paper argues that individuals do not recognize the effect on future tastes/preferences or being able to account for the future shadow price of smoking. In Suranovic et al (1999) this diverse perception of the individuals ability to consider the future effects of present behavior is justified by the assumption of the widespread knowledge of the health effects of smoking. The empirical implication of the reasoning is that there will not be any variables of future consumption or price in the estimated models.

## 3.3 Health

In Grossman's human capital approach, health is a durable whose output is healthy days. Individuals demand health capital due to both the consumption and the investment perspective. First, individuals enjoy being healthy, meaning that health yields utility directly. Secondly, individuals can allocate the healthy days at the market, which yields monetary returns, or in the production of different commodities. According to Grossman's theory, health differs to other human capital since it affects the time constraint rather than the productivity. However, other scholars (e.g. Behrman & Deolalikar 1988) argue the opposite. When considering the approach of this paper, the difference is semantic and has no competing implications.

Analogously with other capital, health depreciates, and individuals therefore need to invest in their health to keep it at the desired level. According to Grossman, it is extremely plausible that sooner or later the depreciation rate will be increasing with age. In other words, more investments are needed and it will be more expensive to keep the capital at the same level.

Health Investments, as Medical Care or exercise, is produced in the same way as other commodities, using purchased goods and own time. Grossman analyzes the effect of an increase in human capital – education – analogously to a technological shift in a production function. The individual becomes a more efficient producer; the health demand curve is therefore shifted upwards. How this effect occurs is unclear (e.g. Akin et al 1985). For instance, followers of Grossman (e.g. Muurinen 1982; van Doorslaer 1987) have discussed the possibility of differences in the ability of efficiently allocating health resources.

Grossman's third main variable, wage, affects both the shadow price and the monetary return of health. In the context of opportunity costs, a higher wage increases the value of healthy time. Simultaneously, own time is an input of the production function of health investments. However, unless health is produced with own time as the only input, wage will be positively correlated with the demand for Health.

Although the models neoclassical assumptions of certainty and full information has been criticized (e.g. Wagstaff 1986 and Zweifel & Breyer (1997)), Grossman's model has much to offer to the theoretical framework of this paper. Reasonably, individuals do not choose the moment of their own death based on the shadow price of health, but plausibly the health behavior of individuals affects the health status at the margin. This is certainly the case regarding smoking, which has been considered to both to affect the depreciation rate and being a negative investment in health (compare Akin et al 1985; Wagstaff 1986). The cost side of the smoking decision – mostly negative effects on health status and the future earnings – will, in the remaining of this thesis, be discussed using Grossman's model as a framework.

## 3.4 Intoxication

As stated, individuals partly smoke to enjoy the physical pleasure of being intoxicated. On the supply side of the model, smoking, and other substances, is, together with own time, the inputs in the production of Intoxication. In similarity to other commodities, the demand for intoxication is affected by income and the price of the inputs. In relation to the discussion about addiction past consumption affects both sides of the equation.

Further, substance use or intoxication is a way of coping with the daily round and handle with stressful moments in life. Hence, a deprivation in quality of life and a stressful environment will have an increasing effect on the demand for intoxication. The coping effect has been extensively discussed in the psychological literature; Wills & Hirky (1995) identify the most important coping functions of substance use (intoxication) as distractions from the problems and a change in the affective state. Furthermore, Chaney & Roszell (1985), Wills (1990), Cohen & Williamsson (1988) all suggest that life-stress increases the risk of substance use. Above all, individuals will turn to substance use if they lack other coping resources. In economic terms, intoxication is then a substitute for more expensive and pleasuring commodities as traveling, luxury food or a tenderly and supporting environment.

To summarize, stressful events in life such as financial stress, unemployment or a death of family member yield a need for coping resources. At the same time, the available resources will vary among individuals due to their characteristics and social environment.

## 3.5 Peer or Social acceptance

As suggested by Emery et al (2001), Powell et al (2005) and DiCecca et al (2000) smoking is a probable input in the production function of social or peer acceptance. Especially among youths, smoking patterns can be a result of peer pressure or a willingness to fit in to certain groups. However, older individuals probably respond to similar incentives. DiCecca et al (2000) argue that the availability of possible substitutes to smoking in the production function of peer acceptance vary among peer groups, which affect the derived demand for smoking. The differences can partly be explained by cultural and socioeconomic factors. Arguably in certain context, especially among adults in higher social groups, smoking could, if it is socially unaccepted, even be a negative investment in Peer Acceptance. Further, Peer Acceptance is a durable, which depreciates over time and therefore a, more or less, continuing investment pattern is required to uphold the desired level. Past investments probably has a similar reinforcement effect as in the context of Intoxication. An individual who has invested – started to smoke – to be socially accepted in a peer group has therefore incentives to keep smoking, since the marginal utility may increase.

# 4 Data & Variables

## 4.1 Overview of the data and study population

The data used in the study is from Statistic Sweden's Survey of Living Condition (ULF), linked to a number of Swedish registers to include more precise measures of variables such as income. Since 1975, Statistic Sweden has annually interviewed randomly selected respondents between age 16-84 (in certain years older individuals have been included) about their living conditions. From 1980 and onwards there has been a special focus on health related issues every eight year. These "health-waves", 1980/81, 1988/89 and 1996/97, are the ones used in this specific study. The response rate, which has been declining over time from 86% in 1980 to 78% in 1997, is comparably high in an international perspective (SCB 2001).

In the waves of interest, around half of the observations are from a rotating panel, providing the possibility of using more advanced panel data methods. To take advantage of the additional information of repeated observations, only individuals in the panel have been used. The population has been defined to a cohort of individuals age 20-68 in 1980/81; i.e. born between 1912 and 1961[3]. This gives the possibility of following the same individuals over the three consecutive waves. The lower limit is due to the decomposition exercise of the second step of this study. Reasonably, income is not a fair indicator of the living condition for individuals below the age of 20. Furthermore, only individuals who were included in the panel in 1980/81 are included in the study.

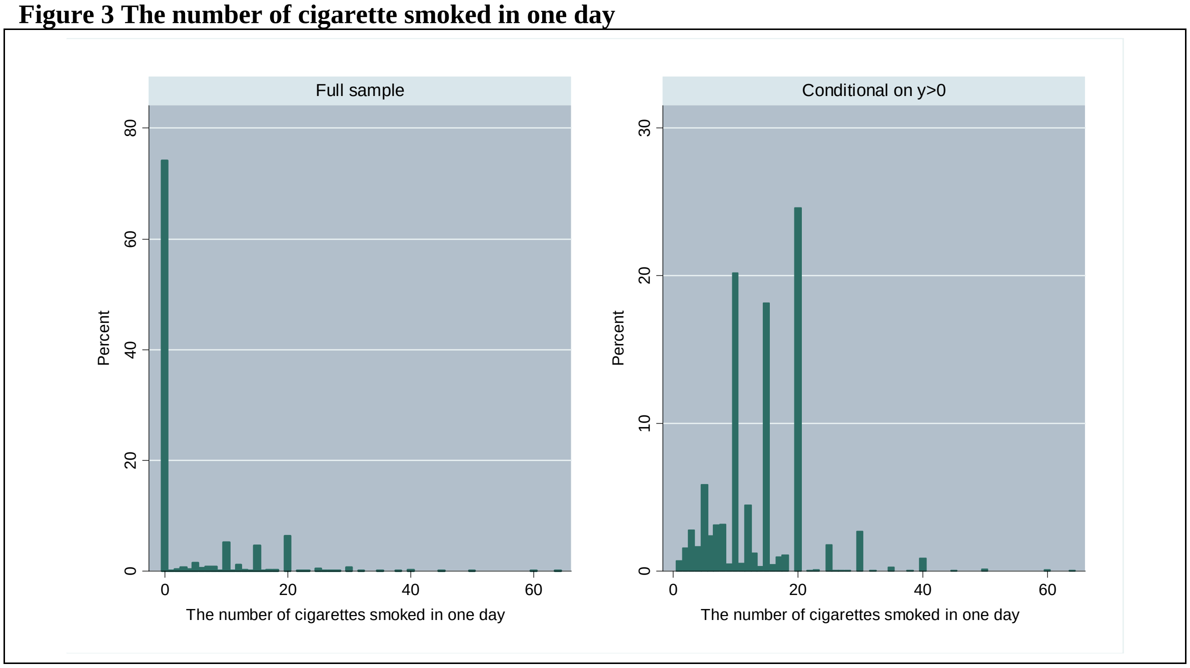
The unbalanced panel includes a sample of 5224 individuals and 13004 total observations. Due to non-response in a number of questions in the questionnaire, the sample is reduced to 5168 individuals and 12703 observations. In the study, two set of models are estimated, where the use of a lag in one set further reduces these numbers to 3992 individuals and 7216 observations.

[3] In the second step of this project, such a definition makes it possible to study health inequalities in an aging population

## 4.2 The dependent variable

The distribution of the dependent variable, the number of cigarettes smoked during one day, is neither continuous nor discrete. As seen in Figure 3 there is a mass zero outcome since around 75% of all observations are non-smokers. Further, the spikes in right diagram of Figure 3, showing all positive observations, clearly illustrate a reporting issue often referred to as heaping. When respondents are asked to report their daily cigarettes consumption, they tend to “round off” the number of cigarettes, reporting fractions or multiplies of a packet rather than an exact number (compare Clark & Etilé 2002).

**Figure 3 The number of cigarette smoked in one day**




Source: Ulf-data, Statistic Sweden
Note: all waves included

However, as long as the latter issue is seen as a measurement error (error-in-variables problem), there is no need for consideration. A random measurement error in the dependent variable does not yield biased estimators (compare Wooldridge 2003, p 302-303). The former issue will be addresses later on.

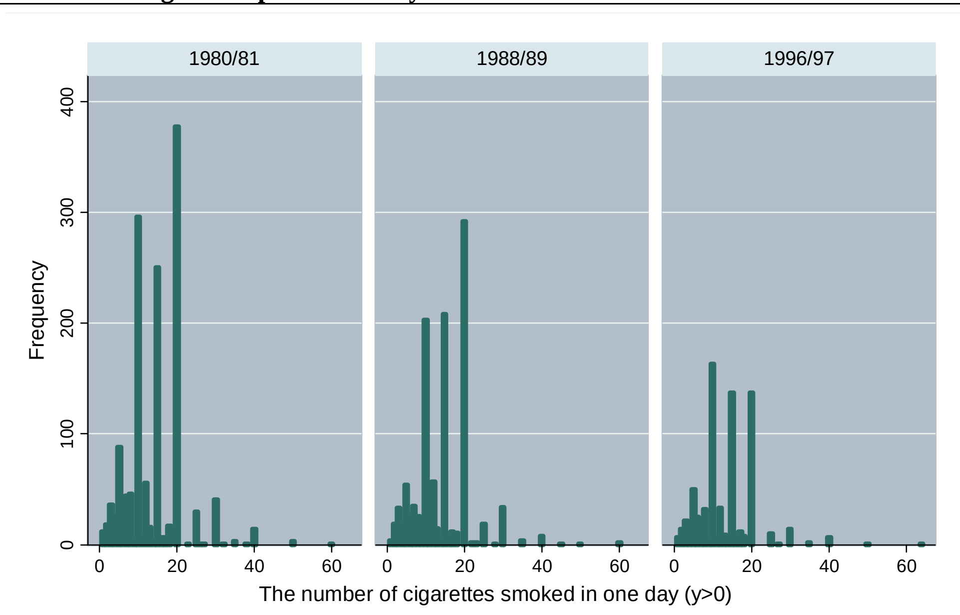
The decline of smoking prevalence over time, discussed in the introductory chapter, is also observed in the studied longitudinal dataset. Even though the mean conditional on positive observations has decreased, the largest part of the decline seems to be due to the amount of smokers rather than the decreased number of cigarettes per smoker. The illustration of the conditional distributions by waves, in

Figure 4, tells the same story. The spikes look smaller but the pattern is consistent.

**Table 1 Descriptive statistics of the dependent variable**

| | | Observation | Mean | St.dev. | Min | Max |
|---|---|---|---|---|---|---|
| ***Wave 1:*** | | | | | | |
| | Total | 4636 | 4.394521 | 7.708965 | 0 | 60 |
| | >0 | 1435(31%) | 14.19721 | 7.267512 | 1 | 60 |
| ***Wave 2:*** | | | | | | |
| | Total | 4286 | 3.702053 | 7.27064 | 0 | 60 |
| | >0 | 1107(26%) | 14.33333 | 7.230886 | 1 | 60 |
| ***Wave 3:*** | | | | | | |
| | Total | 3781 | 2.543771 | 6.044325 | 0 | 64 |
| | >0 | 735(19%) | 13.08571 | 7.071607 | 1 | 64 |

**Figure 4 Number of cigarettes per smoker by waves**




Source: Ulf-data, Statistic Sweden

## 4.3 The Independent variables

### *4.3.1 Addiction*

As discussed in the theoretical framework, past consumption affects the smoking decisions. The data provides two opportunities to measure the addictive effect of past consumption on the number of cigarettes: Using the number of years the individual has smoked or a lag of the dependent variable. In a cross-sectional setting Yen & Jones (1996) measure addiction as the maximum number of cigarettes smoked in one day, the number of years one smoked could be an analogous approach. However, such an approach would not exploit the dynamic advantages of the longitudinal data (compare e.g. Jones & Labeaga 2003; Labeaga 1993, 1999; Chaloupka 1991, Becker, Murphy & Grossman 1994; DiCecca et al 2008)

Since the panel only includes three waves, using the number of years as a measure of past consumption instead of a lag increases the number of observations by (around) 50%. It also avoids endogeneity problems of a dynamic model. On the other hand, a dynamic model provides the opportunity to more realistically model the addiction and state dependency. Instead of accepting the costs of any of the two approaches, this paper will present and discuss results from both models, one including the lag of the dependent variable and another including the number of years one smoked as a measure of past consumption.

To distinguish between myopic and rational individuals a lead of consumption is normally included in a rational addiction framework. Since the lead, just as the lag, of the dependent variable obviously is endogenous, an instrument is needed for identification. Normally, future prices of cigarettes are used. Following Gilleskie & Strumpf (2005) and Clark & Etilé (2002), no lead is included in this study. Prices are not known by certainty, especially not when it is eight years between the waves. It would be impossible for individuals to perfectly foresee prices and tax changes within such a large timeframe (compare Gilleskie & Strumpf 2005). In addition, since the price variable available is the yearly mean of the consumer price index, there is little variation between the individuals in each wave (compare Yen & Jones 1996). This is in line with the theoretical framework developed in Chapter 3, rejecting the idea of individuals capable of superhuman calculations. It is also outside the scope of this paper to conduct a test of the rational addiction hypothesis.

### 4.3.2 Education and Socio economic status

In terms of Grossman's model higher ***education*** increases the demand for health, since health can be more efficiently produced (Grossman 1972). If this raise in productivity is derived from increased health knowledge, this would arguably decrease the demand for intoxication, or at least promote substitution to less addictive and health damaging inputs. In a sense this would be a more efficient allocation of resources. Moreover, it is reasonable to believe that the demand for or inputs to peer acceptance vary between different social environments, which in part is a function of one's level of education, profession and ***socio-economic group***. In certain contexts, smoking possibly even has a negative effect in the production function. In more general terms, there may be less of acceptance to smokers among higher social groups. To capture this effect in the econometric model, eight dummies indicating the self reported socio-economic group have been included. The definitions are in accordance with the definitions provided by Statistic Sweden and employed in earlier research in the area (SCB 1982; Lindström et al 2000; Nystedt 2006). The first five - *Unskilled manual workers* (***sei1***), *Skilled manual workers* (***sei2***), *Assistant non-manual employees* (***sei3***) and *Intermediate non-manual employees* (***sei4***), *Employed and self-employed professionals, higher civil servants and executives* (***sei5***) - are considered to be ordinal ranked. This provides a possibility to observe and analyze a socioeconomic gradient (compare Lindström et al 2000; Nystedt 2006). In contrast, *self-employed entrepreneurs in different scales* (***sei6)*** are a very heterogeneous group including both individuals who run their own small family restaurant and owners of medium sized stock companies. The last two groups include individuals involved in the *agricultural sector* (***sei7***) and *students* (***sei8***).

Since the justification of the variable is based on the reasoning of the social acceptance of the surrounding environment, it would not be appropriate to merge the heterogeneous retired individuals into one social group, more or less creating an additional age dummy. Instead, retired individuals are included in the socioeconomic group they belonged to during their working life. The same logic applies to unemployed[4] and home working individuals, as well as the military servants.

### *4.3.3 Income & financial stress*

The shadow price of smoking includes the value of healthy time; the possible future earnings. A higher income (wage) increases the value of healthy time and therefore also the economic incentives to not involve in risky health behavior such as smoking. On the demand side a higher income also provides more coping resources. In the case of finding oneself in a stressful situation, a higher income provides other possibilities to cope than turning to intoxication and smoking.

[4] Although unemployment was considered as variable indicating stress in similarity with financial stress, it was excluding since it did not provide any additional explaining power.

Following Kamrul et al (2007), the income measure, defined as full income, consists of the annual disposable income and the annuity of net wealth added together. This inherited definition will, for consistency with other studies, be used as the ranking variable in the future decomposition exercise. Disposable income consists of income from capital, employment and transfers net of taxes. The data of taxable net wealth of the household from the National Income Tax Statistics has been converted to annuities at a market value. Moreover, the amounts from the different years have been converted into a common scale using the Consumer Price Index. The OECD equivalent scale has also been employed to obtain net income per adult in the household (see Kamrul et al 2007 or Gerdtham & Johannesson 2002 for more detailed information).

To capture the economic and social stress of not being able to support oneself, a dummy indicating the experience of ***financial stress*** has been included. Respondents who have been forced to borrow or ask for social assistance to pay the rent or to buy food during the last year have been coded to ones (compare Lindström 2004; 2007; Nystedt 2006).

### *4.3.4 Demographics*

According to Grossman's model, individuals, when ***aging***, need to invest more to uphold the health capital at a given level, since the capital depreciates at a higher rate (Grossman 1972). As health becomes more expensive, the demand for health decreases while the demand for health investments probably increases. Reasonably, absence from smoking can be considered as a health investment (compare Wagstaff 1986; Gerdtham et al 1999, Akin et al 1985). As the social environment varies among age groups, so does the demand for peer acceptance. In particular, smoking initiation is an adolescent phenomenon, older individuals tend to quit rather than initiate smoking (compare Foster & Jones 2001; Lundborg & Lindgren 2004).

Grossman and his apprentices discuss the possibility that ***sex*** and ***ethnicity*** affect the productivity of health, the depreciation rate (compare Wagstaff 1986; Muurinen 1982). Both factors probably also affect the demand for peer acceptance as well as the inputs included in the production function. Powel et al (2005) discuss peer effects among adolescents and show cultural or ethnical differences to be important determinants of smoking patterns. To catch the ethnicity effect, dummies for first generation immigrants is included in the analysis[5].

Due to specialization and increased return of scale, a shared household yields a higher real income and more resources available (compare Becker 1965). Cohabiting or being married, in contrast to living ***alone***[6], therefore has an income effect that can provide additional coping resources other than intoxication. To have a companion reasonably also increases the alternative coping resources and thereby decreases the demand for intoxication.

Having ***children*** living in the household arguably force individuals to internalize the externalities of smoking, since their children are exposed to second hand smoking. To control for varying social environment due to geographical heterogeneity, a set of regional dummies has been included. The three largest cities – Stockholm, Gothenburg and Malmö – have been merged into a ***City*** dummy, while the remaining of the country has been divided into a ***Southern*** and ***Northern*** region[7].

[5] First generation is defined after country of birth. Even though this obviously is a time-invariant, the value of the individual varies over time. To attain time-consistency within the individuals, values have been adjusted to be the same over the three periods. This has been performed through a rule of majority, to reducing the impact of miscoding. In cases when individuals only are observed twice, the initial value has been used. A dummy indicating second generation was considered, but excluded. The limited number of observations created both statistical and practical issues.

[6] In Sweden the matrimony in itself, even though it gives certain economic security, does not provide similar economic incentives to specialization as in other countries. Cohabiting and married is therefore treated equally and combined into one base group (compare Smith 2002).

[7] The ***Northern*** region is defined as Norrland, while the Southern contains ***Svealand*** and ***Götaland.***

**Table 2 Descriptive statistics; Independent variables**

| *Variable* | *Defintion* | *Model:* | *Static Obs=12703 Ind= 5168* | | *Dynamic =7216 Ind = 3992* | | *Static/Dynamic* | |
|---|---|---|---|---|---|---|---|---|
| | | | **Mean** | **St dev** | **Mean** | **St dev** | **Min** | **Max** |
| **Education:** | | | | | | | | |
| **Educ1** | =1 if Primary education | Omitted | | | | | | |
| **Educ2** | = 1 if Secondary education 2yrs (2 yrs Gymnasium) | | .3148075 | .4644575 | .3197062 | .466395 | 0 | 1 |
| **Educ3** | = 1 if Secondary education 3 yrs (3 yrs Gymnasium) | | .0933638 | .290953 | .0936807 | .2914042 | 0 | 1 |
| **Educ4** | =1 if University education | | .2167992 | .4120809 | .2333703 | .4230052 | 0 | 1 |
| **Sex** | | | | | | | | |
| **female** | = 1 if female | omitted | | | | | | |
| **male** | = 1 if male | | .4902779 | .4999251 | .483786 | .4997717 | 0 | 1 |
| **Children** | | | | | | | | |
| **Childd1** | = 1 if no children in the household | Omitted | | | | | | |
| **Childd2** | =1 if 1 child in the household | | .1450838 | .3521992 | .126663 | .3326181 | 0 | 1 |
| **Childd3** | =1 if 2 children in the household | | .156341 | .3631926 | .1322062 | .3387383 | 0 | 1 |
| **Childd4** | = 1 if >2 children in the household | | .0612454 | .2397893 | .0605599 | .2385377 | 0 | 1 |
| **"Marital Status"** | | | | | | | | |
| **Married/Cohabiting** | = 1 if Married or Cohabiting | omitted | | | | | | |
| **alone** | = 1 if single household | | .2631662 | .4403692 | .2573448 | .4372013 | 0 | 1 |
| **Ethnical** | | | | | | | | |
| **Im1** | = 1 if born in Sweden (or having at least one Swedish parent) | omitted | | | | | | |
| **Im2** | = 1 if First generation immigrant | | .0736047 | .2611367 | .0676275 | .251123 | 0 | 1 |
| **Socioec Status** | | | | | | | | |
| **sei1** | = 1 if Unskilled manual worker | omitted | | | | | | |
| **sei2** | = 1 if Skilled manual worker | | .1557112 | .3625957 | .1492517 | .3563611 | 0 | 1 |
| **sei3** | = 1 if Assistant non-manual employee | | .1527198 | .3597314 | .1543792 | .3613369 | 0 | 1 |
| **sei4** | = 1 if intermediate non-manual employee | | .1573644 | .3641583 | .170316 | .3759362 | 0 | 1 |
| **sei5** | = 1 if Employed and self-employed professional, higher civil servant and executive | | .1157207 | .3199023 | .1281874 | .3343215 | 0 | 1 |
| **Sei6** | = 1 self-employed entrepreneur | | .0820279 | .274418 | .089939 | .2861142 | 0 | 1 |
| **sei7** | = 1 if employed in agricultural sector | | .0335354 | .180037 | .0357539 | .1856888 | 0 | 1 |
| **sei8** | =1 if student (or unclassified military servants) | | .0175549 | .131332 | .0072062 | .0845888 | 0 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Age:* | | | | | | | | |
| **Age1** | = 1 if 10-25 | omitted | | | | | | |
| **Age2** | = 1 if 26-45 | | .2345903 | .4237591 | .2351718 | .4241356 | 0 | 1 |
| **Age3** | = 1 if 46-55 | | .2159332 | .4114844 | .2422395 | .4284681 | 0 | 1 |
| **Age4** | = 1 if 55-65 | | .1770448 | .3817216 | .1787694 | .3831857 | 0 | 1 |
| **Age5** | = 1 if 66-75 | | .1398882 | .3468847 | .1787694 | .3831857 | 0 | 1 |
| **Age6** | = 1 if 76-84 | | .0621113 | .2413671 | .0979767 | .2973038 | 0 | 1 |
| **Regions** | | | | | | | | |
| **Cities** | =1 if Sthlm, Gothenburg or Malmö | omitted | | | | | | |
| **Northern** | = 1 if southern Sweden | | .1954656 | .3965743 | .2001109 | .4001108 | 0 | 1 |
| **Southern** | = 1 if Northern Sweden | | .5146028 | .4998064 | .5277162 | .4992658 | 0 | 1 |
| **lninc** | = natural log of full income | | 11.78298 | .5193756 | 11.84197 | .5179649 | .458225 | 15.7773/ 15.1656 |
| **Financestress** | = 1 if financial stress, if one had to borrow or get social assistance to afford food or housing within the last year | | .0972999 | .2963773 | .0879989 | .2833129 | 0 | 1 |
| **Time dummies** | | | | | | | | |
| **Wave1** | = 1 if 1980/81 | omitted | | | | | | |
| **Wave2** | =1 if 1988/89 | | .3374006 | .4728415 | omitted | | | |
| **Wave3** | = 1 if 1996/97 | | .2976462 | .4572411 | .4492794 | .4974552 | 0 | 1 |
| **Smokepast** | = nr years smoked | | 10.99591 | 13.84184 | | | 0 | 70 |
| **smokelag** | | | | | 3.928492 | 7.393176 | 0 | 60 |
| **pastlag** | = nr of years smoked in the last period | | | | 9.900637 | 12.56255 | 0 | 60 |
| **ydlag** | = 1 if smoking in last period | | | | .2778548 | .4479724 | 0 | 1 |
| **inyd** | = 1 if smoking in initial time period | | | | .2973947 | .4571434 | 0 | 1 |
| **mage** | = individual mean of age | | 49.1924 | 13.73841 | 49.27979 | 13.59243 | 20/24 | 84/76 |
| **mlninc** | = individual mean of ln(income) | | 11.78146 | .4062381 | 11.79057 | .374885 | .710991/ 6.95183 | 15.66252/ 13.8722 |

# 5 Econometric Methods & Empirical Modeling

The econometrical discussion in this chapter starts out in a cross-sectional setting to account for the problem of excess zeroes in the dependent variable. Arguments for choosing the 2PM model above the Sample Selection model is presented, before moving on to the panel data setting. The Mundlak-type specification is discussed as an attempt to parameterize the individual heterogeneity. A panel and a pooled version of both the static and the dynamic model are employed to obtain more robust results. Since computing the Random Effect Probit is time consuming the pooled versions are needed to present reliable bootstrapped standard errors of the full model. The justification of using Generalized Linear Models and a heteroskedasticity robust retransformation are elaborated respectively. Furthermore, the solutions to the problem of the endogenous initial state in the dynamic models are discussed. Finally, the computations of the expectations, and the marginal and incremental effects of the 2PM are shortly presented.

## 5.1 Econometric methods

### *5.1.1 A hurdle approach*

Following the standard approach of estimating smoking behavior, individuals engage in two decisions; whether to smoke and the amount of cigarettes. Such a model consists of two equations; one for participation and one of consumption. If the latent variable justification of the binary model is applied this is normally called a hurdle model (Cameron & Trivedi 2005; Jones 1989). As worked out by Jones (1989) the more precise specification of the model depends on two crucial assumptions; the issue of dominance and the degree of independence between the disturbance terms of the two equations.

If the disturbance terms are assumed to be independent, the hurdle model of Jones (1989) is reduced to the Cragg-model (see Cragg 1971). If we instead assume first hurdle dominance, i.e. that the participation decision dominates the decisions of intensity, the model is identical to Heckman's Sample Selection model (Heckman 1979). In the smoking context first hurdle dominance implies that the individual first decides whether to smoke and then considers the quantity. Combining the two assumptions, first hurdle dominance and random independent disturbance terms, further reducing the model and gives the following log likelihood (compare Jones 1989; Madden 2008).

$$\ln L = \sum_{y=0} \ln\left[1 - \mathrm{P}(y > 0 \mid x)\right] + \sum_{y>0} \left\{\ln\left[\mathrm{P}(y > 0 \mid x)\right] + \ln\left[g(x \mid x, y > 0)\right]\right\} =$$

$$\left\{\sum_{y=0} \ln\left[1 - \mathrm{P}(y > 0 \mid x)\right] + \sum_{y>0} \ln\left[\mathrm{P}(y > 0 \mid x)\right]\right\} + \left\{\sum \ln\left[g(x \mid x, y > 0\right]\right\}$$

$$= \ln L_1 + \ln L_2$$

The log likelihood above shows that the two parts, participation and conditional consumption, can be estimated separately. Hence, the model is called a 2 Part Model (2PM). In this application the first part is a Probit, while the second part uses the continuous dependent variable.

### *5.1.2 On the choice of the 2PM vs. the Sample Selection Model*

In contrast to studies using data of recorded cigarette consumption (see Labeaga 1999; Jones & Labeaga 2003) respondents are asked to report their typical consumption pattern. Since zero outcomes represent non-smokers, and do not arise from infrequency of purchases, first hurdle dominance can be assumed (compare Jones 1989; Madden 2008). Consequently, there is a choice between the 2PM and the Sample Selection model. A choice that been widely discussed in the literature of health econometrics (see Jones 2000, Dow & Norton 2003, Madden 2008). Heckman (1979) shows that the sample selection model can be estimated in two steps as well. The main difference between the computations of the models is the inclusion of the Inverse Mills Ratio (IMR)[8] in the conditional equation of the Sample Selection model.

Following the criteria highlighted in Dow & Norton (2003) the 2PM is preferred above the Sample Selection model. First, the observed zeros in the data are actual outcomes, generated from non-smokers. Among others, Dow & Norton (2003) argue that the 2PM is more appropriate to compute actual marginal effects than the sample selection model. The latter model was rather developed to analyze potential outcome, a latent variable that is not fully observed. In the classical example, female labor participation, actual wage is not the outcome of interest. Instead, due to the self selection in participation, the potential wage bid, if one were actually participating,

[8] The Inverse Mills ratio is defined as $\lambda = \phi(X\beta)/\Phi(XB)$, the standard normal PDF divided by the standard normal CDF of the Probit (Wooldridge 2002 p 522).

is the latent variable of interest. Although it is possible to compute actual Marginal Effects in the Sample Selection Model, it is more complicated than normally presented in the literature (Dow & Norton 2003).

Secondly, to be fully empirically identified the Sample Selection model needs an exclusion restriction; i.e. there has to be at least one variable in the participation equation that do not appear in the second. Otherwise the inclusion of IMR in the second equation risks causing a problem of collinearity, since IMR is a function of the independent variables (see Verbeek 2004 p 232). As in other health economic applications, such a variable is not available in the data (compare Seshamani & Gray 2004). The assumption of independence between the errors can be statistically tested. However, due to the collinearity problem discussed above, such a test has been criticized (Dow & Norton 2003).

### 5.1.3 A Panel data setting

The longitudinal data makes it possible to control for unobserved heterogeneity. These unobserved individual effects, $\alpha$, are included in the Error Component Model; in a general formula expressed as;

$$y_{it} = \beta' x_{it} + \alpha_i + e_{it}, \quad i = 1,2....,N,\ t = 1,2....,T,$$

where $x_{it}$ and $\beta$ are $k \times 1$ vectors. This notation will be used further on in the chapter.

A set of assumptions decides the appropriateness of the Random Effect (RE) or the Fixed Effect (FE) estimator respectively. The former estimator requires stronger assumptions[9], while the latter excludes the possibility of using time invariant variables. A third alternative – the pooled model – ignores $\alpha$ completely and is estimated as a cross-sectional model clustering on the individuals. As other Quasi Maximum Likelihood Estimators only the first moment, the mean, needs to be correctly specified for consistency (see Wooldridge 2002; Jones 2007).

The panel version of the 2PM requires not only independence between the errors, but also independence between the unobserved individual heterogeneity, $\alpha$, of the two equations. Van Ourti (2004) and Tooze et al (2002) solves the problem by estimating the two parts of the

[9] RE assume no correlation between the individual effect and error term, while FE (or within-) estimator is consistent under weaker assumptions. Still the FE can not distinguish the observed heterogeneity from the unobserved. For a more comprehensive overview see Wooldridge (2002) or Cameron & Trivedi (2005).

Random Effect 2PM simultaneously using Maximum Likelihood. Such an approach complicates computations and interpretations of the marginal effects, and is therefore not employed in this paper.

In other applications of the Random Effect 2PM, independence of the individual effects is assumed. Seshamani & Gray (2004) compute both a Random Effect and a pooled version of the separate 2PM without discussing the problem. In a paper about cocaine demand, Grossman et al (1998) use the separate 2PM in a rational addiction approach. In the first step a Linear Probability Model (LPM) is used and a linear regression in the second. The presence of endogenous lags and leads of the dependent variable is solved by IV-techniques. None of the two applications derives the marginal effects, although they both use the assumption of separate probabilities in either deriving predictions of the expected value or elasticities.

In the first equation, the Probit, a consistent Fixed Effect estimator is not available. Furthermore, the RE of the second equation is rejected in both a classical and an alternative Hausman test. However, many of the variables of interest, when modeling smoking behavior, are, more or less, time invariant. These variables are also needed for the future aims of this study; the decomposition of a Concentration Index. Following Jones et al (2007 p 217) a different solution is therefore implemented.

To handle the problem of unobserved heterogeneity a Mundlak type specification is used (Mundlak 1978, Wooldridge 2002, Jones et al 2007). In a similar way to what Chamberlain (1980; 1984) suggests, Mundlak (1978) attempts to parameterize the individual effect of the Random Effect-model using the within-groups mean of the independent variables.

$$\alpha_i = \alpha' \bar{x}_i + u_i, u \sim iidN(0, \sigma^2)$$

Since the model is estimated conditional on the individual effects, this allows for correlation between the effects and the independent variables. In practice, this is done by augmenting the model by adding a vector of $\bar{x}_i$ to the equation. In this application, the mean of the continuous variables ***age***[10] and ***ln(income)*** is used. The approach will work in the pooled models as well,

[10] Although age is included as a set of dummy variables, the mean of the continuous variable is used in the Mundlak specification.

and for simplicity reason the Mundlak type specification of the two steps are presented as pooled models.

$$P(y > 0 \mid x) = \Phi\left[(\beta_1' x_{it} + \alpha_1' \overline{x}_i\right]$$

where $\Phi(\cdot)$ is thestandard normal CDF.

$$\ln y_{it} = \alpha_2' \overline{x}_i + \beta_2' x_{it} + e_{it}, e \sim iidN(0, \sigma^2)$$

Using the Mundlak approach, instead of the Fixed Effect estimator, to handle the individual heterogeneity allows for including the time invariant variables, sex (***male***) and ethnicity (***im2***) in $\overline{x}_i$. The error term from the Mundlak equation ($u_i$) is now included in the error term of the augmented equation ($e_i$), which is assumed to be independent and identically distributed. Since the individual effect is parameterized – leaving only white noise in the error term – the argument for estimating the 2PM in two separate stages is now stronger.

Implementing the 2PM in a panel setting introduces an additional issue in the second equation (see Ahn 2004). Since the conditional population (smokers) varies over time, this launches an unbalanced panel with gaps. According to Ahn (2004), inference has to be made with cautions. However, in the data there are only 42 individuals who quit smoking between the first and the second wave, and then pick up smoking in the third. Hence, the issue shall not be of importance.

## 5.2 Empirical modeling

### *5.2.1 Further delimitations*

Due to the trade-off between exploiting the information of a dynamic model and including a larger amount of observations, both the number of years of past smoking and a lag of consumption will be used in separate models to model addiction. The models are myopic in the sense that only past, and not future, consumption is considered. However, as argued in the presentation of the theoretical framework, this still allows individuals to consider future health effects.

The dynamic setting, i.e. the inclusion of a lag of the dependent, does not only provide possibilities, it also introduces a problem of endogeneity in both the Probit and the second equation. Addiction can be thought of as state dependency, where the initial condition will be endogenous (compare Gilleskie & Strumpf 2005; Jones et al 2007 p 250). Simply put, the first observation of an individuals smoking status ( $y_{i1}^{d}$ ) is not the true initial value ( $y_{i0}^{d}$ ), but a function of the earlier history that plausibly will be correlated with the unobserved heterogeneity.

The Probit is preferred above other binary models as the Logit and the LPM. The former is asymptotically similar to the Probit, however the Probit appears as standard in the 2 PM. In the LPM the predictions, which are used in the derivation of the marginal effects, are not guaranteed to be inside the interval of [0-1] (compare Kennedy 2003 259-261)[11]. For a comprehensive explanation of the Random Effect Probit (see Jones 2007 or Wooldridge 2002)

In applications of the 2 Part Model it has become standard to use the log of the dependent variable in the second equation to attain a less skewed and a more appropriate distribution (see Afifi et al 2007; Seshamani & Gray 2004; Manning & Mullahy 2001). Since the expected values now will be in logs, a retransformation back to levels is needed for a meaningful interpretation of the marginal effects. However, taking the exponential of the expected value is not sufficient. If the errors are assumed to be homoskedastic, the smearing factor of Duan (1983) can be employed. An alternative to OLS with a logged dependent variable suggested by Manning & Mullahy (2001) is the Generalized Linear Models with a log link, which is not as sensitive to heteroskedasticity[12]. Using (Quasi) Maximum Likelihood GLM fits models as follows;

$$g\{E(y \mid x)\} = \beta' x \,, \quad y \sim F$$

where $g(\cdot)$ is the link function, which is used for the retransformation back to level, and F is the assumed distribution of the errors. If $g(\cdot)$ is the natural log function and F is assumed to be the gamma[13] distribution the following model applies (see Statacorp 2007).

$$\ln\{E(y \mid x)\} = \beta' x \,, \quad y \sim gamma$$

Since heteroskedasticity is present, this is performed in the pooled versions of the two models. The software used for the analysis, Stata, does not provide a Random Effect Estimator of the

[11] If predictions outside the interval are forced to 0 and 1, predictions with certainty are obtained although the occurrence of the reverse is still possible.

[12] In difference to the simple regression of ln(y) on x, $E(y) = e^{x\beta}$ holds.

[13] A Modified Park test did not suggest a specific distibution. The result, was a compromise between the gamma and the poison distribution. Gamma seems to be standard in the applied application and is therefore assumed.

Generalized Linear Models[14] (see Statacorp 2007). Instead a linear regression of ln(y) is run and a heteroskedastic robust retransformation suggested by Baser (2007) is performed to attain the coefficients in levels (see Appendix A for technical details).

### *5.2.2 Empirical specification of the four models*

To clarify how the econometric methods have been employed, the following part will present the empirical modeling of first the static and then the dynamic model.

In the static model, first the Probit is run as;

$$\Pr(y_{it}^{d} = 1) = \Phi(\beta' x_{it} + \delta z_{it} + \alpha' M_{i} + \chi' t^{S}),$$

$y_{it}^{d}$ = binary variable for smoking(0-1) ; M: mean(age), mean(ln(income)), male, im2

$x =$ Education (Educ2-4), Regions(South, North), ln(income), alone, financial stress, Socio-economic group (SEI2-8), Children(Childd2-4), Age(Age2-6), and a constant term

$t^{S}$ = time period in the static models(wave2-3)

Where, M is a vector of the Mundlak-variables to parameterize the individual effect including the mean of both age and ln(income), and time-invariants. Further z, is the indicator of addiction ***smokepast***, i.e. the number of years the respondent have been smoking. Finally, x is a vector of the remaining independent (exogenous) variables. In the second step the same variables are used in the regression conditional on positive outcomes.

$$\ln y = \beta' x_{it} + \delta z_{it} + \alpha' M_{i} + \chi' t^{S}$$

Due to the issues arising in the dynamic specification, additional variables are included in the dynamic models. Since the solutions to the endogenous initial condition vary between the linear and the non-linear models, the variables differ between the two parts. Instead of z, number of years smoked (***smokepast***), the lag of the dependent is used as an indicator of addiction or state dependency. Similar to Grossman et al (1998) the lag of the binary dependent $y^{d}$ is included in the Probit;

$$\Pr(y_{it}^{d} = 0) = \Phi(\beta' x_{it} + \eta^{d} y_{i,t-1}^{d} + \alpha' M_{i} + \gamma y_{i1}^{d} + \chi t^{d})$$

, $y_{t-1}^{d}$ = lag of $y^{d}$ , $y_{1}^{d}$ = initial smoking state (0-1), $t^{d}$ = time period in the dynamic models(wave3)

and the lag of the continuous y is included in the second equation.

[14] The Stata command – xtgee – implies the Population Average estimator instead.

$$\ln y = \beta' x_{it} + \eta y_{i,t-1} + \alpha' M_i + \chi t^d ,$$

To solve the endogeneity problem of the initial condition in the Probit $y_1^d$, the smoking state in the initial time period, is included as an additional variable to parameterize the individual effects as suggested by Wooldridge (2000; 2005). Since the Wooldridge-approach is developed for non-linear models, instrumental variable (IV) is carried out as a solution to the endogeneity problem in the second equation.

In many applications (e.g. Becker et al 1994, Grossman et al 1998), lagged prices are used as instruments. In the lack of a varying price variable, another instrument needs to be identified. Since the number of years the respondent has smoked, $z$, is used as an indicator of addiction in the static model, $z_{i,t-1}$ can be considered a possible instrument for $y_{i,t-1}$. Reasonably, the number of years one smoked before the last period affects smoking decisions in that period only, implying that $z_{t-1}$ has no direct effect on y. Decisions in the next period are only indirectly affected through the state of addiction, i.e. the number of cigarettes smoked in the last period. Simply put $z_{t-1}$ affects $y_{t-1}$, that in turn has an effect on $y$. Hence, $z_{t-1}$ satisfies the conditions of an appropriate instrument, being correlated with the endogenous variable ($E(z_{i,t-1} \mid y_{i,t-1}) \neq 0$), but uncorrelated with the residuals ($E(z_{i,t-1} \mid e_{it}) = 0$) (compare Cameron & Trivedi 2005 p 95-101). Other possible instruments, justified with similar reasoning, are the socio-economic group of the respondent's father and lags of other independent variables. However, these variables appear to either fail a test for weak instruments or cause a rejection of an overidentification test.

In the pooled version the IV solution is carried out in two stages. In the first step the reduced form equation, $y_{t-1} = \beta'_{RF} x_{it} + \delta z_{i,t-1}$, is regressed. Secondly, the predictions from the reduced form are included in the GLM equation.

$$E(y \mid x) = \exp(\beta' x_{it} + \eta \hat{y}_{i,t-1} + \gamma M_i + \chi t^d)$$

Normally, the standard errors are invalid after the two stages. However, since the bootstrapped standard errors are obtained normal inference is possible. In the panel version, the IV procedure is performed directly through an inbuilt Stata command –xtivreg- which performs a two-stage least square generalization of the random effect estimator (Statacorp 2007).

### 5.2.3 *Marginal effects and bootstrapped standard errors*

The expected value, $E(y \mid x)$, of the 2 Part Model is in a general formula expressed as;

$$E(y \mid x) = \Pr(y > 0 \mid x) * E(y \mid x, y > 0)$$

where $\Pr(y > 0 \mid x)$ is the probability of a positive outcome and $E(y \mid x, y > 0)$ is the expected value conditional on a positive outcome.

The marginal effect of a continuous variable can be derived by using the product rule of derivatives;

$$\frac{\partial E(y \mid x)}{\partial x} = \Pr(y > 0 \mid x) * \frac{\partial E(y \mid x, y > 0)}{\partial x} + E(y \mid x, y > 0) * \frac{\partial \Pr(y > 0 \mid x)}{\partial x}$$

while the incremental effect of the dummy variables can be defined as

$$\begin{aligned} &E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0) \text{ [15]} \\ &\quad = \left\langle \Pr(y > 0 \mid x, x_d = 1) - \Pr(y > 0 \mid x, x_d = 0) \right\rangle * E(y \mid x, x_d = 0).. \\ &\quad ..- \Pr(y > 0 \mid x, x_d = 0) * \left\langle E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0) \right\rangle \end{aligned}$$

Expectations, marginal and incremental effects for the two parts and the full model specifically derived for this application is presented in more detail in Appendix A. The Marginal Effects can be computed at the mean of all observations or as the Average of the Marginal Effects (AME) of all observations (See Cameron & Trivedi 2005 p.122, 467; Bartus 2005). Calculating the Marginal Effects at the mean implies that individuals will be assigned implausible values, therefore the Average Marginal Effects is preferable. The Average Marginal Effects are computed for all the pooled versions of the models. The Total Marginal Effect and the Marginal Effect of the Probit are calculated as an average of the total population, while the Marginal Effect (levels) in the second equation only uses the subsample with positive outcomes.

Because the residuals from the conditional equation are used in the heteroskedasticity robust retransformation, calculating the Total Average Marginal Effect of the full sample is complicated. Instead, the Marginal Effects at the mean are presented. Analogously to the pooled versions, the Total and Probit Marginal effects are estimated at the mean of the full sample, and

[15] The expression is as presented in a presentation by Partha Deb, Willard Manning and Edward Norton at the ASHE (American Society of health Economics) 2006 conference. Analogously, the incremental effects in the two steps are defined as $\Pr(y > 0 \mid x, x_d = 1) - \Pr(y > 0 \mid x, x_d = 0)$ and $E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0)$ respectively.

the second equation levels at the mean of the subsample. The same logic applies to the incremental effects of the dummies.

The standard errors of the retransformation and the 2PM can be calculated by the delta method, however due to all the derivatives that would be tedious. Instead, bootstrapped standard errors are presented (compare Ai & Norton 2000). Due to the longitudinal setting of the data, the bootstrap procedure is performed with clustering on each individual (compare Seshamani & Gray 2004). Instead of randomly drawing, with replacement, subsample of observations as in a cross-sectional data, the full procedure is replicated 999 times by subsamples of individuals.

However, due to time consuming computation of the RE Probit, standard errors are obtained by the delta method, and the retransformation of the second equation is bootstrapped individually. Moreover, the presented estimates of bootstrapped standard errors of the Total Marginal Effects of the panel version are only replicated 99 times, and shall therefore be interpreted with caution (compare Statacorp 2007).

# 6 Results & Discussion

## 6.1 Diagnostics

### *6.1.1 Attrition and non-response bias*

Survey Panels often suffer from attrition and non-response. In an overview, Nicoletti & Peracchi (2005) point out a number of reasons to non-response: Death, movement out of the survey sample, e.g. emigration, refusal to respond, absence of the person at the address, and other types of non-contact. The ULF-data includes information of the occurrence of death between the waves but not about other reasons for attrition. If we believe smoking or some of the independent variables affect the non-response, rather than that individuals are missing at random, the non-response or attrition is non-ignorable (see Jones 2007 p 265-272).

In an article about attrition in health surveys Jones et al (2006) suggest a test proposed by Verbeek & Nijmans (1992). To test for non-response bias a variable reflecting the attrition is included in the original model. If the respondents are missing at random such a variable (r) shall not be correlated with y when all the independent variables are held constant; e.g. the null tested is $E(y|x,r)=E(y|x)$. Three sets of models are estimated, using three different indicators (r); a count for the number of waves the respondent appears in and two separate dummies – one for whether the respondent appears in all waves and one indicating appearance in the next wave.

Since, in general, the null hypothesis tested is $E(y|x,r)=E(y|x)$, in a two part setting reasonably $(Pr(y>0|x)*E(y|x,x>0)= (Pr(y>0|x,r)*E(y|x,r,x>0)$ is the null under consideration. If the data was cross-sectional such a test could have been performed by an LR-test. However, the pooled or panel setting introduces a problem of dependence between the observations making the test inadequate. Instead, two sets of tests are performed using the expectations of the dependent variable in two equations separately. The variable indicating attrition, r, is then tested by a Wald test. The inclusion of a dummy indicating appearance in the next wave reduces the sample of the dynamic model to just including one wave. Hence, a test using panel methods is not possible.

**Table 3 Verbeek-Nijman test for non-response bias**

| $H_0$ *Indicators* | Pooled | | Panel | |
|---|---|---|---|---|
| | Static | Dynamic | Static | Dynamic |
| Pr(y>0\|x)=Pr(y>0\|x,r) | | | | |
| **r1= Nr of waves** | **0.3164** | **0.7372** | **0.0763** | **0.6168** |
| **r2=All waves** | **0.3134** | **0.7372** | **0.0817** | **0.6168** |
| **r3=Next wave** | **0.5651** | **0.2663** | **0.5651** | **NONE** |
| E(y\|x, x>0)=E(y\|x,r,x>0) | | | | |
| **r1= Nr of waves** | **0.4711** | **0.4177** | **0.1881** | **0,4135** |
| **r2=All waves** | **0.0996** | **0.7346** | **0.0354** | **0.7294** |
| **r3=Next wave** | **0.8525** | **0.6537** | **0.6325** | **NONE** |

*Note: NONE implies that the model can not be tested, ether due to the specification of the null or the r variable. Next wave" reduces the number of waves to 1 in the dynamic model, and the LR-test is not applicable to the panel-setting. The test of r1in the first equation of the dynamic model is equivalent to the test of r2.*

Out of the 22 tests in Table 3, only one rejects the null at a 5% level and three others are close to rejection with p-values between 0.05-0.1. In all, there is little support for the presence of non-response bias. It follows that the models can be estimated without further correction.

### 6.1.2 Pooled versions

Both normality and homoskedasticity of the Probits are rejected in a set of conditional moment tests, performed by regressing a vector of ones on the score contribution (compare Verbeek 2005 p 200-201). In difference to other models, heteroskedasticity in a Probit implies not only incorrect standard errors, but also the wrong functional form and inconsistency. However, the heteroskedasticy is probably produced by either the omission of a variable included in the actual data generating process or the unobserved heterogeneity. In applied work, when trying to model reality with data such a problem is difficult to avoid.

Although $\beta$ is inconsistently estimated if the assumed functional form is wrong, the Probit model may still provide good estimates of the marginal and incremental effects. Wooldridge (2002, p 479) stresses that this is a functional form issue and should be treated thereafter. It is not really of interest to consistently estimate $\beta$ if $P(y=1|x) \neq \Phi(x\beta)$. Even though the Probit is heteroskedastic, it is still preferred due to the disadvantages of its alternatives. First, a Probit adjusted for heteroskedasticity is in general considered to be an instable model and it normally produces inconsistent estimators (see Keele, L. J. and Park, D. K. 2004). Second, the Linear Probability Model provides predictions outside the interval of [0-1], which are used in the derivation of the Total Marginal Effects.

As tests for functional form[16], neither the RESET-test performed as recommended by Jones (2006 p 14-20) nor the Pregibon's Link test reject the null of no misspecification in the static pooled Probit. In contrast, the dynamic Probit fails both tests. The RESET test is performed on the second equations GLM as well. The null is rejected in the static model, but not in the dynamic. For consistency in an instrumental variable context, the RESET test is performed using the Pagan-Hall(1983)/Pesaran-Taylor(1999) statistics; i.e. not using predictions of $X\hat{\beta}$ but instead $Z\hat{\beta}$ or $\hat{X}\hat{\beta}$ [17](compare Baum et al 2003; 2007). Both models are considered to be heteroskedastic. However, since the employed retransformation is insensitive to heteroskedasticity, this is not an issue. Furthermore, the two pooled Probits as well as the full pooled 2PM pass Copas-test[18] for overfitting, whereas the second equations Generalized Linear Models fail.

### *6.1.3 Panel versions*

Since deriving the score vector in the RE Probit is extremely complicated (if possible), there is no common test for heteroskedasticity. However, both the static and the dynamic model fail a conventional RESET-test. An alternative Hausman test rejects the RE-estimator (compare Jones et al 2007 p 215-219); following the earlier discussion this suggests a Mundlak type specification. Further, the static model of the conditional equation passes the RESET test while the dynamic fails. For consistency in the IV context the Pagan-Hall (1983) statistics is computed, using the predictions from a random effect regression of ln(y) on all the exogenous dependent variables and the instruments (compare Baum et al 2003; 2007).

---

[16] The RESET test for probit-models, recommended by Jones (2006 p 14-20), includes a square on the prediction of the latent variable (xb) in a auxiliary probit model. Further, the Pregibon Link test is performed by testing $x\beta$^2 in a model also including $x\beta$ . The RESET is at times referred to as a test for Omitted Variables bias, the command –ovtest- in Stata performs such a test (Statacorp 2007). Wooldridge (2002 124-125) criticizes such an approach, arguing that the RESET-test only test for the functional form and for the omissions of higher powers of the dependent variables.

[17] where Z is a vector of the instruments including the exogenous variables as instruments for themselves

[18] Copas test for overfitting in a one way version randomly splits the sample into two groups. The coefficients from the first model $(\hat{\beta}_1)$ are used to retain a forecast of the second sample $(y_2 = x_2\hat{\beta}_1)$ . The original y of the second sample is then regressed on its forecast. This procedure is repeated 999 times to test whether the coefficient of the forecast is equal to one.

To not mix up the error term with the unobserved heterogeneity, the Bickel-version of the Breusch-Pagan test for heteroskedasticity is performed by an auxiliary regression using the error term from the Fixed Effect estimators. As before, The Pagan-Hall statistics is used in the IV context. Both tests reject the null of homoskedastic errors. However the residuals are less heteroskedastic and have a less skewed distribution when using the log of the dependent variable instead of y. The estimation output of the Random effect models, as seen in the Appendix B, show that more than 50% of the unexplained variation is due to within individual variation. The ***rho***>0.5 in all but one model implies that the Random Effect models tell an additional story of the data in comparison with the pooled models.

### 6.1.4 Instrumental variables

The lag of the indicator of addiction used in the static model, years smoked, is employed as an instrument for the endogenous lagged dependent variable. In a test for weak instruments using the reduced form equation, the null is rejected. An Omitted Variable version of the Hausman-test[19] rejects the null of exogeneity; in a regression of all the independent variables, both endogenous and exogenous, the instrument ***pastlag*** shows up to be significant in both the pooled (p-value=0,0692) and the panel (p-value=0.0064) version of the test (compare Kennedy 2003 p 173). Although the p-value is larger than 0.05 in the pooled model, conservative inference of the diagnostics suggests an instrumental variable approach to be employed in both models.

**Table 4 Misspecification Tests**

| | Static | | Dynamic | |
|---|---|---|---|---|
| | **Pooled** | **Panel** | **Pooled** | **Panel** |
| **Cond. moment tests** | ***Fail*** | | ***Fail*** | |
| **RESET Probit** | ***Pass*** | ***Fail*** | ***Fail*** | ***Fail*** |
| **RESET second eq** | ***Fail*** | ***Pass*** | ***Pass*** | ***Pass*** |
| **Pregibon Probit** | ***Pass*** | | ***Fail*** | |
| **Copas Probit** | ***Pass*** | | ***Pass*** | |
| **Copas GLM** | ***Fail*** | | ***Fail*** | |
| **Copas 2PM** | ***Pass*** | | ***Pass*** | |

### 6.1.5 Implications on inference

Since the models fail some of the misspecification tests, the results shall be interpreted with

[19] The matrix of the classical test was not positive definite, which is not an unusual problem when there is only one endogenous variable. (compare Kennedy 2003 p 173)

caution. The models appear to not fully capture the unobserved heterogeneity, which probably causes the misspecification problems. In part, this is due to the restrictions set up by the further aims of the study, excluding both interaction terms and other possibilities to model the heterogeneity. Still, the models are considered preferable to the alternatives.

None of the models can be argued to be preferable above the others, since the tests all implies different conclusions. Nevertheless, as seen in following part of the chapter, the results are similar throughout the models, indicating that the results still capture a picture of the reality.

## 6.2 Results

In the following part of the chapter are the results from the four models presented and discussed.

**Table 5 Probit Marginal & Incremental Effects**

| | Static | | | | Dynamic | | | |
|---|---|---|---|---|---|---|---|---|
| Model | Pooled *Obs=12003* | | Panel *Obs=12003* | | Pooled *Obs=7216* | | Panel *Obs=7216* | |
| Variables | Coef | p>z | Coef | p>z | Coef | p>z | Coef | p>z |
| **Smokepast** | .0172936 | *0.000* | .0099824 | *0.000* | - | - | - | - |
| **Ydlag** | - | - | - | - | .4232073 | *0.000* | .1062983 | *0.024* |
| **Inyd** | - | - | - | | .1353993 | *0.000* | .4591535 | *0.000* |
| **Mage** | -.008643 | *0.000* | -.0050003 | *0.000* | -.0019202 | *0.009* | -.0019153 | *0.006* |
| **Mlninc** | -.0082903 | *0.380* | -.002079 | *0.765* | -.0075706 | *0.622* | -.0094325 | *0.425* |
| **Male** | -.1072936 | *0.000* | -.0661595 | *0.000* | -.0102956 | *0.149* | -.008153 | *0.233* |
| **im2** | .0371608 | *0.005* | .0296469 | *0.012* | .0012994 | *0.922* | .0015057 | *0.904* |
| **wave2** | -.0673614 | *0.000* | -.0392117 | *0.000* | - | - | - | - |
| **wave3** | -.1605535 | *0.000* | -.0859818 | *0.000* | -.0436712 | *0.000* | -.0491516 | *0.000* |
| | | | | | | | | |
| **childd2** | -.0029791 | *0.736* | -.0037708 | *0.458* | -.0170165 | *0.129* | -.0153069 | *0.039* |
| **childd3** | -.0288585 | *0.002* | -.0166799 | *0.000* | -.0355427 | *0.004* | -.0248601 | *0.002* |
| **childd4** | -.0526067 | *0.000* | -.0238787 | *0.000* | -.0514847 | *0.001* | -.0324078 | *0.000* |
| **educ2** | -.0011854 | *0.885* | -.0008288 | *0.866* | 9.32e-06 | *0.999* | -.0023506 | *0.757* |
| **educ3** | -.0201473 | *0.101* | -.0145562 | *0.007* | -.0245482 | *0.067* | -.0198431 | *0.022* |
| **educ4** | -.0541446 | *0.000* | -.0281923 | *0.000* | -.0466661 | *0.000* | -.0338869 | *0.001* |
| | | | | | | | | |
| **Alone** | .0537436 | *0.000* | .0375471 | *0.000* | .0212916 | *0.014* | .0213638 | *0.022* |
| **age2** | .0114572 | *0.283* | .0040467 | *0.520* | .0305788 | *0.039* | .0230257 | *0.077* |
| **age3** | -.0193799 | *0.232* | -.0136126 | *0.072* | .0283079 | *0.087* | .0216468 | *0.196* |
| **age4** | -.0766086 | *0.000* | -.034074 | *0.000* | .0164146 | *0.441* | .0160231 | *0.462* |
| **age5** | -.1272794 | *0.000* | -.0460614 | *0.000* | -.0085466 | *0.719* | -.0069276 | *0.724* |
| **age6** | -.1570106 | *0.000* | -.0428625 | *0.000* | -.0010089 | *0.972* | -.0062738 | *0.783* |
| | | | | | | | | |
| **sei2** | -.0218512 | *0.017* | -.0075195 | *0.141* | -.0048914 | *0.646* | -.0016045 | *0.861* |
| **sei3** | -.0121313 | *0.231* | -.0010217 | *0.863* | .007355 | *0.515* | .0099062 | *0.350* |
| **sei4** | -.0136517 | *0.208* | -.0076495 | *0.194* | -.0150657 | *0.232* | -.0121479 | *0.188* |
| **sei5** | .0141107 | *0.314* | .011349 | *0.239* | .0187373 | *0.221* | .0162419 | *0.302* |
| **sei6** | -.0173553 | *0.206* | -.0034249 | *0.641* | -.0092719 | *0.473* | -.0035567 | *0.748* |
| **sei7** | -.0781054 | *0.008* | -.0237838 | *0.003* | -.0467354 | *0.047* | -.0272745 | *0.025* |
| **sei8** | -.0467549 | *0.005* | -.0216603 | *0.001* | .0019545 | *0.960* | .0099339 | *0.783* |
| | | | | | | | | |
| **Lninc** | -.0082903 | *0.380* | -.0053689 | *0.217* | -.0146022 | *0.244* | -.0115996 | *0.092* |
| **Finance stress** | .0408573 | *0.000* | .0320268 | *0.000* | .0395816 | *0.001* | .0366348 | *0.024* |
| **Southern** | -.012826 | *0.133* | -.0106295 | *0.035* | -.0004017 | *0.961* | -.0030438 | *0.692* |
| **Northern** | .0080721 | *0.445* | .0038758 | *0.557* | .0023596 | *0.823* | .0009782 | *0.920* |

Note; The pooled versions are presented as AME, average of the individual marginal effects while the panel versions are presented as Marginal Effects at the mean (which is considerably smaller). P-values from pooled versions are from bootstrapped standard errors, while the dynamic are from standard errors derived using the delta method (as done by –margeff- command in Stata)

**Table 6 Second Equation Marginal & Incremental Effects**

| | Static | | | | Dynamic | | | |
|---|---|---|---|---|---|---|---|---|
| Model | Pooled GLM *Obs=3277* | | Panel Retransformed lny *Obs=3277* | | Pooled GLM *Obs=1638* | | Panel Retransformed lny *Obs=1638* | |
| variables | Coef | p>z | Coef | p>z | Coef | p>z | Coef | p>z |
| **smokepast** | .1766397 | *0.000* | .137428 | *0.000* | - | - | - | - |
| **Smokelag** | - | - | - | - | .5722195 | *0.000* | .6607916 | *0.000* |
| **mlninc** | .0178665 | *0.963* | -.0173354 | *0.959* | -.2014261 | *0.840* | -.5614872 | *0.279* |
| **mage** | -.1352509 | *0.000* | -.137673 | *0.000* | -.1052145 | *0.050* | -.0855801 | *0.103* |
| **male** | 1.978484 | *0.000* | 1.295675 | *0.000* | 2.67716 | *0.000* | .6781006 | *0.072* |
| **im2** | 1.570752 | *0.004* | 1.496424 | *0.006* | .6629131 | *0.311* | -.0722008 | *0.898* |
| **wave2** | -.530636 | *0.086* | -.7890434 | *0.002* | - | - | - | - |
| **wave3** | -2.435162 | *0.000* | -2.878126 | *0.000* | -1.545519 | *0.001* | -1.53168 | *0.001* |
| **childd2** | .1146074 | *0.757* | -.2686806 | *0.374* | .0217668 | *0.963* | -.3878508 | *0.383* |
| **childd3** | .300339 | *0.445* | .1022282 | *0.743* | .5873858 | *0.333* | -.177927 | *0.738* |
| **childd4** | .8395598 | *0.175* | .3557148 | *0.496* | 1.361214 | *0.148* | -.2590122 | *0.761* |
| **educ2** | -.2271623 | *0.466* | -.1606693 | *0.550* | -.1546257 | *0.696* | -.0972023 | *0.791* |
| **educ3** | -.160002 | *0.759* | -.1060009 | *0.821* | -.2880588 | *0.694* | -.6633778 | *0.299* |
| **educ4** | -1.643818 | *0.001* | -1.227594 | *0.006* | -.845143 | *0.206* | -.7663012 | *0.198* |
| **alone** | .9977029 | *0.002* | 1.084613 | *0.000* | .4398154 | *0.276* | .7158689 | *0.059* |
| **age2** | .5901906 | *0.215* | .5001183 | *0.219* | .9035475 | *0.205* | .5493689 | *0.430* |
| **age3** | .595985 | *0.424* | .5120897 | *0.441* | 1.718215 | *0.118* | .6216507 | *0.536* |
| **age4** | -1.099204 | *0.249* | -.5458403 | *0.532* | 1.099663 | *0.490* | .1918793 | *0.887* |
| **age5** | -2.605989 | *0.026* | -1.523502 | *0.149* | -.2001452 | *0.916* | -.5657043 | *0.726* |
| **age6** | -3.459697 | *0.007* | -2.594102 | *0.008* | -1.607041 | *0.416* | -1.220248 | *0.483* |
| **sei2** | -.7628181 | *0.042* | -.3266029 | *0.296* | -.5013257 | *0.301* | -.4380341 | *0.364* |
| **sei3** | .1146845 | *0.785* | -.1928024 | *0.610* | .2914424 | *0.577* | .0145626 | *0.976* |
| **sei4** | .4337238 | *0.386* | .2505445 | *0.501* | 1.087905 | *0.104* | .3153191 | *0.561* |
| **sei5** | .593166 | *0.384* | -.2033701 | *0.708* | .350015 | *0.649* | .0198088 | *0.978* |
| **sei6** | 1.244222 | *0.046* | .7862301 | *0.107* | 1.2775 | *0.123* | .7172384 | *0.279* |
| **sei7** | -5.153272 | *0.000* | -3.827901 | *0.038* | -6.251677 | *0.000* | -4.864879 | *0.000* |
| **sei8** | -1.053983 | *0.152* | -.4682512 | *0.475* | -.0967875 | *0.948* | -.2859077 | *0.876* |
| **lninc** | .0178665 | *0.963* | -.0173354 | *0.959* | .3939469 | *0.486* | -.5614872 | *0.279* |
| **finance stress** | 1.573306 | *0.000* | .8721447 | *0.006* | 1.264396 | *0.014* | .9140568 | *0.085* |
| **southern** | -.7016669 | *0.058* | -.6048584 | *0.051* | -.5223571 | *0.262* | .0800095 | *0.840* |
| **northern** | -1.308712 | *0.002* | -.6193628 | *0.109* | -1.155382 | *0.029* | -.1240406 | *0.796* |

Note; The pooled versions are presented as AME, average of the individual marginal effects while the panel version are presented as Marginal Effects at the mean. P-value from bootstrapped standard error.

**Table 7 Total Marginal & Incremental Effects**

| | Static | | | | Dynamic | | | |
|---|---|---|---|---|---|---|---|---|
| Model | Pooled *Obs=12003* | | Panel *Obs=12003* | | Pooled *Obs=7216* | | Panel *Obs=7216* | |
| variables | Coef | p>z | Coef | p>z | Coef | p>z | Coef | p>z |
| **smokelag** | | | | | .1271657 | *0.000* | .014381 | *0.002* |
| **smokepast** | .2723241 | *0.000* | .0936381 | *0.000* | | | | |
| **Ydlag** | | | | | 4.998433 | *0.000* | .7665274 | *0.024* |
| **Inyd** | | | | | 1.665703 | *0.000* | 3.296288 | *0.000* |
| **Mlninc** | -.1043004 | *0.392* | -.0027326 | *0.967* | -.1399578 | *0.584* | -.0388652 | *0.721* |
| **Mage** | -.1481244 | *0.000* | -.0490031 | *0.000* | -.0475267 | *0.001* | -.0156418 | *0.004* |
| **Male** | -.6917585 | *0.000* | -.5769865 | *0.000* | .4886251 | *0.000* | -.0417778 | *0.400* |
| **im2** | .8772297 | *0.000* | .3497639 | *0.006* | .163513 | *0.451* | .0089638 | *0.913* |
| **wave2** | -1.047351 | *0.000* | -.3914347 | *0.000* | | | | |
| **wave3** | -2.955938 | *0.000* | -.8851019 | *0.000* | -.9487468 | *0.000* | -.3942355 | *0.000* |
| | | | | | | | | |
| **child2** | -.0096875 | *0.948* | -.0391295 | *0.401* | -.2101953 | *0.209* | -.11723 | *0.037* |
| **child3** | -.3004835 | *0.058* | -.1548101 | *0.000* | -.3160101 | *0.108* | -.1815262 | *0.004* |
| **child4** | -.4770519 | *0.040* | -.221808 | *0.000* | -.3490764 | *0.229* | -.2354869 | *0.002* |
| **educ2** | -.0738955 | *0.585* | -.0116847 | *0.802* | -.0342452 | *0.798* | -.0192741 | *0.750* |
| **educ3** | -.3072941 | *0.144* | -.1394816 | *0.012* | -.3768581 | *0.111* | -.153112 | *0.018* |
| **educ4** | -1.173855 | *0.000* | -.2862456 | *0.000* | -.7976788 | *0.001* | -.2561 | *0.002* |
| | | | | | | | | |
| **Alone** | .9305746 | *0.000* | .3988863 | *0.000* | .3601895 | *0.006* | .176607 | *0.015* |
| **age2** | .2977203 | *0.104* | .0550777 | *0.329* | .5680241 | *0.011* | .1839308 | *0.097* |
| **age3** | -.0979337 | *0.735* | -.1138881 | *0.138* | .7119591 | *0.016* | .1759709 | *0.191* |
| **age4** | -1.317659 | *0.001* | -.3268319 | *0.000* | .445076 | *0.297* | .1213164 | *0.453* |
| **age5** | -2.424397 | *0.000* | -.4457968 | *0.000* | -.152448 | *0.770* | -.0625833 | *0.688* |
| **age6** | -3.037945 | *0.000* | -.4135567 | *0.000* | -.3697739 | *0.501* | -.0721553 | *0.619* |
| | | | | | | | | |
| **sei2** | -.4890231 | *0.002* | -.0796439 | *0.091* | -.1737602 | *0.286* | -.0222473 | *0.731* |
| **sei3** | -.1298198 | *0.468* | -.0112893 | *0.857* | .1564267 | *0.380* | .0716721 | *0.345* |
| **sei4** | -.0670098 | *0.733* | -.0649449 | *0.271* | .0560706 | *0.795* | -.081041 | *0.282* |
| **sei5** | .3351714 | *0.186* | .1071917 | *0.230* | .3106436 | *0.204* | .1175325 | *0.305* |
| **sei6** | .0929673 | *0.697* | -.0077254 | *0.916* | .1689113 | *0.474* | -.0089575 | *0.905* |
| **sei7** | -2.372171 | *0.000* | -.2760167 | *0.000* | -1.99446 | *0.000* | -.2458411 | *0.002* |
| **sei8** | -.892661 | *0.002* | -.2112483 | *0.000* | .0030507 | *0.996* | .0624821 | *0.833* |
| | | | | | | | | |
| **Lninc** | -.1043004 | *0.392* | -.0489475 | *0.333* | -.096064 | *0.538* | -.0957035 | *0.237* |
| **finance stress** | .9193194 | *0.000* | .3459324 | *0.001* | .7611087 | *0.000* | .3040179 | *0.016* |
| **southern** | -.356608 | *0.024* | -.1157674 | *0.010* | -.1213356 | *0.403* | -.0198638 | *0.726* |
| **northern** | -.2251589 | *0.213* | .0157554 | *0.806* | -.2261324 | *0.185* | .0038934 | *0.954* |

Note; The pooled versions are presented as AME, average of the individual marginal effects while the panel version are presented as Marginal Effects at the mean. Due to the time consuming computation, the p-value of the panel versions are based on bootstrapped standard errors with 99 replications only.

### 6.2.1 *Addiction*

Irrespective of the variable used to model addiction, it is shown to be an important determinant of smoking patterns. This is in line with earlier literature. Both Yen & Jones (1996), using the maximum number of cigarettes smoked per day, and other scholars, using the lag of the dependent, supports the findings (Chaloupka 1991, Labeaga 1999, Jones & Labeaga 2003, Gilleskie & Strumpf 2005). The result is also as expected in terms of the theoretical framework. Higher consumption in the last period increases the demand for intoxication in this period, while the marginal productivity of each cigarette decreases.

The results in the static models imply that, for every year smoked an individual is on average 1-2 percentage points more prone to smoke and consumes, unconditional on smoking, 0.25 more cigarettes per day. The conditional estimate is around 0.15 per year. The effect may at first seem small, but for a difference in 10 years the effect is substantial.

The significance and magnitude of the coefficients on ***ydlag*** ($y_{t-1}^{d}$) – the lag of the dummy indicating smoking in the last period – highlight, in similarity with Gilleskie & Strumpf (2005), the importance of state dependency in smoking behavior. In the context of a dynamic model, Gilleskie & Strumpf (2005) stress the difficulties to distinguish the state dependency from the individual heterogeneity. Given that the Mundlak type specification capture the individual effect, being a smoker increases the consumption by nearly five cigarettes a day. Further, every cigarette in the last period additionally increases the consumption by around 1/8 cigarettes. However, the magnitude of the coefficient on ***ydlag*** ($y_{t-1}^{d}$) appears to vary between the pooled and the panel version and the reverse pattern emerges for to initial state variable ***inyd*** ($y_{1}^{d}$). For some computational reason, the endogeneity of the initial condition seems to capture a larger part of the effect of state dependency in the panel model[20].

### 6.2.2 *Education & Socio-economic status*

As predicted in the theoretical framework, higher education affects both propensity to smoke and the total consumption. The coefficients in the second equation are only significant in the static

[20] This is not an effect of the differences between evaluating the incremental effect at the mean or the average incremental effect.

model. There is also considerable support for an effect of having a longer secondary education (3 years gymnasium) on the propensity to smoke, compared to the base-group of primary education. Still, the Total Incremental Effect is insignificant in all models. Since education is included as a factor affecting the demand or productivity of all the commodities discussed in the theoretical framework, it is problematic to distinguish the effects. Reasonably, a higher education reduces the demand for intoxication through the increased health knowledge and removes cigarettes from the production of social or peer acceptance.

The results are in line with earlier findings; e.g. Hersch (2000) finds education to be significant in the participation as well as the conditional consumption equation for both men and women. In a British study, Clark & Etilé (2002) shows education to be positively correlated with quitting. Since most of the smokers initiated smoking in their youth this is consistent with the negative effect on participation. Kenkel (1991) shows that education still has an effect on smoking behavior even when health knowledge is controlled for. This supports the conclusion that education does not only affect the smoking behavior through the demand of health, but also due to the differences in social acceptance.

Although most of the coefficients of the dummies indicating socioeconomic status are insignificant, the variables are in several of the models jointly significant. Hence, exclusion would possibly cause omitted variable bias. The social gradient analyzed in Nystedt (2006) is shown to be insignificant when estimated conditional on the other independent variables, including education.

Statistic Sweden's definition of the socio-economic group is correlated with educational attainment. Probably, an extensive part of the social gradient and the effect of the social environment are captured by the educational effect. Arguably, the first two socio-economic groups can not be separated by education. In line with that reasoning, the pooled static model suggests that non-manual workers are less prone to smoke, and that they smoke fewer cigarettes than the base-group, manual workers.

In relation to these results, Lindström et al (2000) find that the socioeconomic gradient loses significance when social participation, which often also is correlated with education, is introduced into a model of smoking cessation. In a British context, Yen & Jones (1996) show that, without including education in the model, individuals from higher social groups tend to smoke less.

Individuals employed in the agricultural sector smoke significantly fewer cigarettes than the base group of manual workers. In relation to this, Tauras (2005) finds that living on a farm significantly reduces smoking participation in the US. A reasonable interpretation is that the social environment in the sector affects the inputs in the production of social acceptance.

Students are still climbing on the educational ladder, which implies higher expected earnings in the future and an increased value of healthy days. Additionally, the inputs in the production of social acceptance are probably similar to individuals with higher education. According to the static models students respond to these incentives, being both less prone to smoke and consuming fewer cigarettes than the base group of manual workers. In contrast, the coefficients in all equations of the dynamic model are highly insignificant.

### *6.2.3 Income & financial stress*

The income effect is consistently insignificant in all the models[21], implying that income neither affects the propensity to smoke nor the amount of cigarettes. Higher income increases the value of healthy time and thereby increases the cost of smoking[22]. Seemingly individuals do not respond to these incentives. It may be the case that the effect of an increased value of healthy time is outweighed by the increased possibility of other health investments. However, it is still surprising that a higher income can not provide alternative coping resources.

The results are somewhat contradictory to earlier studies. Hersh (2000) finds a negative effect in both participation and consumption, whereas in Tauras (2005) income affects participation only. In contrast, Raptou et al (2005) finds a positive income effect using Greek data.

---

[21] It is of importance to state that the insignificant income effect is not an effect of the inclusion of financial stress dummy. Alternative models have been tested and even without financial stress income shows up insignificant. Instead the income effect seems to be eaten up by the inclusion of wave-dummies. Increased income tends to be highly correlated by age and time.

[22] This is a simplification: The opportunity cost of time, the wage, is not directly included in the model.

Although the mean of ***ln(income)*** is individually insignificant, all the variables included to parameterize the individual effect in a Mundlak type specification are jointly significant[23]. This indicates that the long term income do not affect the smoking behavior, however, the variable may still capture a part of the individual effect as a confounding variable.

Even though income does not affect the smoking decisions in a clear direction, being under financial stress do. This is one of the most consistent findings in the study. Whereas income affects the budget constrain and provides substitutes to smoking, the effect of experiencing financial stress increases the demand for intoxication. Obviously, resources are not only scarce but also insufficient, making more expensive coping resources unattainable. Except for a p-value of 0.085 in the dynamic panel version of second equation, the coefficients of the dummy are strongly significant in all models. The findings imply that experiencing financial stress affects both the propensity to smoke and the intensity in cigarette consumption.

These results confirm earlier findings in the Swedish context (see Nystedt 2006; Lindström 2004). In an international perspective, similar results are found. Several authors highlight the importance of financial stress as a determinant of relapse and failure of cessation (see Siahpush & Carlin 2006, Dorsett & Marsh 1998; Graham 1993). In line with this study, the authors conclude that it is the stress of not coping with the daily life that affects the smoking decisions.

### *6.2.4 Age*

Although the significance of the age dummies varies, the pattern of an inverse u-curve is consistent throughout the models. The pattern is less clear in the second equation, implying that ageing first and foremost affects the decision to smoke rather than the quantity. The declining propensity and the decreasing intensity of consumption when aging are supported primarily in the static models. The weaker effect in the dynamic model is probably due to the fewer time periods and the variable used to model the addiction. In contrast to the consistent support of the inverse u-curve, there are considerable differences in the magnitude of the coefficients between the two sets of models. The pattern of the inverse u-curve is confirmed by earlier studies. In a study of

[23] The joint significance have been tested in the separate original models.

Irish women in Madden (2008) the pattern appear in the second equation only, while Hersch(2000) finds a similar pattern in both equations for women as well as for men.

Moreover, the findings are consistent with the idea that initiation is a youth phenomenon. When aging, individuals quit smoking or reduce the number of cigarettes. Additionally, individuals appear to respond to the higher cost of smoking in terms of health. Aging makes health investments more expensive; more inputs are needed to produce the same amount of healthy days. Simultaneously, the health capital depreciates with a higher rate. Thus, more investments are needed to outweigh the negative effect of smoking.

As discussed, the Mundlak type specification tries to parameterize the unobserved individual heterogeneity by including the mean of age. Given the assumption that individuals are affected by their social environment during their youth, the mean of age capture a generational effect, caused by the development of the society and changes in social values over time. The demand of social acceptance and the inputs in the production functions probably vary between the generations. Since the smoking prevalence in Sweden has varied over time, as discussed in the background chapter, such a generational effect plausibly exists. Given that the individual mean of age capture and hold this generational effect constant, the support for interpreting the u-curve pattern of the age dummies as an effect of aging alone increases.

### *6.2.5 Sex*

The coefficients on the male dummy in the participation equations imply that men are less prone to smoke. Nevertheless, conditional on smoking, men consume more cigarettes per day. The total unconditional effect is unclear, since the results of the static and the dynamic models are contradictable. The signs in the separate equations are consistent with the 2PM of Raptou et al (2005) using a Greek dataset. Although the results are at odds with the American study in Pacula (1997), the reasoning behind the results can be applied. In Sweden, men, instead of women, tend to mix their tobacco consumption between different goods. Hence, the pattern of the signs can probably be explained by the extensive male snus-use. For men, snus is reasonably a better substitute than for women, especially as an input in the production of social acceptance. By looking at the diagrams in Chapter 1, men seem more responsive to price changes. Given the assumption that the changes in the consumer price index of tobacco reflects the price change of cigarettes rather than of snus, the difference in price sensitivity between the sexes can be explained by the substitute available; i.e. snus.

### *6.2.6 Ethnicity*

Although the sign of nearly all of the coefficients on ***im2***, first-generation's immigrants, are positive, the statistical significance differs between the models. The pooled model implies that, everything else equal, first generation immigrants are more prone to smoke and have a higher intensity in their consumption of cigarettes. A plausible interpretation is that the stress of living in a foreign culture may increase the demand for intoxication. As discussed above, an alternative explanation is that the different social environment and cultural values affect the inputs in the production of peer acceptance. The findings are consistent with both Nystedt (2006) and Lindström & Sundquist (2002). However, the dynamic model provides no support for such a hypothesis.

A reasonable interpretation of the results in the two set of models are that much of the behavior may be explained by the lag of the dependent dummy variable, since smoking is initiated in the youth many of these individuals may have been smoking when arriving to Sweden. Although the endogeneity of the initial smoking condition have been corrected for, state dependency may be hard to distinguish from the individual heterogeneity. This reasoning advocates the explanation of differences in inputs in the production function of peer acceptance rather than the increased demand for intoxication. The findings of Powell et al (2005) provide further support, showing that youths from minorities are smoking less - not more - than their majority counterparts, which rules out the stress of living in a majority culture as an explanation.

### *6.2.7 Household variables*

Living alone compared to the base group of either being married or cohabiting, increases both the probability to participate and the intensity of the conditional consumption. The positive total incremental effect is also statistically significant. The results highlight that living alone decreases the coping resources available. This is probably a result of both the lack of a companion and the less efficient household production in the absence of specialization. The findings are supported by Nystedt (2006), who shows that having a partner has a negative effect on smoking. Further, Clarke & Etilé (2002), Ross & Chaloupka (1999), and Madden (2008) all find marriage to positively affect quitting probabilities.

Seemingly, parents internalize the cost of their children's health; having children in the household reduces the propensity to smoke. The coefficient of the dummies of having two (childd3) and >2 (childd4) children in the household significantly differ from zero in all of the four Probit models. The effect seems to increase with the number of children. In contrast, no support is given for a similar effect in the second equation. Concludingly, the increased shadow price affects the participation decision only; if an individual has chosen to smoke, the number of children in the household does not affect the quantity. The total incremental effect is negative, but only significant in the static model. In a study of Irish women, Madden (2008) finds a competing pattern. The number of children appears insignificant in the participation equation, but positive significant in the second equation of both a 2PM and a Sample Selection model.

### *6.2.8 Regional dummies*
The regional dummies are overall insignificant in the participation equation, whereas the second equation supports, primarily in the static versions, that everything else equal smokers in the larger cities – Stockholm, Gothenburg and Malmö – have a higher intensity in their consumption. Although the support is weak for an overall effect, the dummy of the Southern region is significant in the pooled models

### *6.2.9 A decreasing time trend*
Unsurprisingly, and consistent with the declining smoking prevalence showed in the descriptive statistics, the propensity to smoke as well as both the unconditional and the conditional consumption decrease over time[24].

[24] It shall though be stated that exclusion of the time dummies change the inference of the income effect. Since the study follows the same cohort, a large proportion of the individuals probably become richer over time. The time trend may therefore capture part of the possible income effect. On the other hand, waves can be seen as dummy for different social values and different cigarettes prices.

# 7 Summary & Concluding remarks

## 7.1 Main findings

The major policy interest of the research project will first and foremost be the future decomposition exercise, where the driving factors behind the inequalities in health risks will be revealed. Still, this study provides an additional understanding of the determinants of smoking.

In short; a theoretical framework based on the household production model has been developed A pooled and a panel version of both a static and a dynamic two part model have been estimated on the Swedish Survey of living conditions (ULF). To handle the unobserved heterogeneity a Mundlak type specification has been used. Irrespectively of the measure of addition – the number of years the respondents smoked or lagged consumption – the smoking decisions is heavily influenced of the passed behavior. Furthermore, conditional on the other dependent variables neither income nor the socioeconomic gradient have an effect on smoking, while the effect of aging appears to be in the shape of an inverse u-curve. Two of the determinants with largest influence, higher education and having children in the household, primarily affect the smoking habits through the participation decision. While higher education is one of the usual suspects recognized as smoking determinants, the number of children is not.

Moreover, the incremental effects of both living in a single household and experiencing financial stress are consistently positive throughout the models. Both the effects highlight that smoking is a probable input in the production of intoxication, which individuals demand for coping with stressful situations. Individuals employed in the farming sector smoke on average considerably less cigarettes compared unskilled workers. In the pooled versions of the models the total incremental effect is as large as -1.99 to -2.37 cigarettes, an effect that primarily comes from the second equation. Among the time-invariant variable, the support for an effect of being a first generation immigrant varies between the static and the dynamic models, while men are less prone to participate but have a higher conditional consumption.

To sum up the most interesting findings; individuals use cigarettes to cope the stress of not being able to support themselves and appear to internalize the cost of their children’s second hand smoking. The findings are overall consistent with earlier literature, implying a high validity of the results. At the same time it is of importance to bear in mind the econometrical issues.

## 7.2 Lessons and thoughts for the future

Empirical modeling is time consuming. Solutions to one issue launch additional problems and extensively increase the work burden. Reasonably, experiencing these problems is a part of the learning process to become a better researcher. Specifically, this study reveals the limitations of parametric methods since several of the tests reject the functional form of the models. The major issue is to capture the unobserved heterogeneity, a common problem in econometrics. This unobservable heterogeneity, in similarity with an omitted variable, causes problems as heteroskedasticity, wrong functional form, and inconsistency.

A lesson learned from the study is that despite all the econometric issues, the models may still produce both plausible and interesting results. There is no need to throw in the towel because of unsatisfying data or incorrect functional forms. Applied research is like solving a maximization problem; given a constraint of time, skills, and available data and methods, one has to maximize the quality of the study. Although a study has flaws, it may still contain essential lessons for policy makers and other readers.

Future improvements could allow for interaction terms[25], which was a restriction of the future decomposition exercise. A more advanced solution to problems of unobserved heterogeneity is a relaxation of the parametric assumptions. E.g. Mixture or Latent Class models as developed in the area of health care by Deb & Trivedi (2002), Bago D´Uva (2006) and Clark et al (2005) could be applied in a smoking context. Such an approach is especially attractive since it is theoretically consistent with the idea of a varying addict-proneness among individuals (compare Orphanides & Zervos 1995; 1998). In the near future an additional wave will become available, which make an extension of the study even more interesting.

[25] It shall though be stated that some experiments to improve the specifications of the models has been conducted without success.

# 8 References

Afifi, Abdelmonem – Jenny Kotlerman - Susan Ettner – Marie Cowan 2007. “Methods for Improving Regression Analysis for Skewed Continuous or Counted Responses”, *Annual Review of Sociology*, 28, p 95-101 Available at SSRN: http://ssrn.com/abstract=1077703

Ahn, Jeonghoon 2004. "Panel Data Sample Selection Model: an Application to Employee Choice of Health Plan Type and Medical Cost Estimation", *Econometric Society 2004 Far Eastern Meetings 560*, Econometric Society.

Ai, Chunrong – Edward C Norton 2000. "Standard errors for the retransformation problem with heteroscedasticity", *Journal of Health Economics*, 19:5, p 697-718

Ai, Chunrong - Edward C Norton 2003. "Interaction terms in logit and probit models", *Economics Letters*, 80:1, p 123-129.

Akerlof, George A. 1991. “Procrastination and Obedience”, *The American Economic Review,* 81:2

Akin, John -CK Griffin - DK Guilkey - BM Popkin 1985. *Demand for Primary Health Services in the Third World*, Rowman & Littlefield.

Auld, M. C. – P. Grootendorst 2004. “An empirical analysis of milk addiction”, *Journal of Health Economics* 23:6, p 1117-1133.

Ault, Richard – T. Randolph Beard – John D. Jackson – Richard P. Saba 2005. "On the (mis)use of cross-price effects to gauge the effectiveness of smokeless tobacco in smoking cessation: A comment on Bask and Melkersson", *European Journal of Health Economics* 50, p 83-86.

Bago D'Uva, Teresa 2006. "Latent class models for health care utilization", *Health Economics* 15. p 329-343.

Bartus, Tamas 2005. “Estimation of marginal effects using margeff”, *Stata Journal* 5:3.

Baser, 0 2007. “Modeling Transformed Health Care Cost with Unknown Heteroskedasticity”, *Applied Economic Research* 01, p 1-6.

Bask, Mikael – Linda Andersson – Maria Melkersson 2003. "Rational Addiction and Cigarette Smoking in the Presence of Bootleg Cigarettes", *Applied Economics Quarterly* 49:4, p 319-338.

Bask, Mikael – Linda Andersson – Maria Melkersson 2006. “Economic Man and the Consumption of Addictive Goods: The Case of Two Goods”, *Substance Use and Misuse*, 41, p 453-466.

Bask, Mikael – Maria Melkersson 2003a “Should One Use Smokeless Tobacco in Smoking Cessation Programs”, *European Journal of Health Economics* 4, p 263-270.

Bask, Mikael – Maria Melkersson 2004 “Rationally addicted to drinking and smoking?”, *Applied Economics*, 36:4, p 373 – 381.

Basu, Anirban -Paul Rathouz 2005 “Estimating marginal and incremental effects on health outcomes using flexible link and variance function models”, *Biostatistics* 6:1 p 93-109.

Baum, C. F. – M. E. Schaffer – S Stillman 2003. “Instrumental variables and GMM: Estimation and testing”, *Stata Journal* 3, p 1-31.

Baum, C. F. – M. E. Schaffer – S Stillman 2007 “Enhanced routines for instrumental variables and GMM: Estimation and testing”. *Working Paper No 667* Boston College Economics

Becker, Gary 1965. “A theory of the allocation of time”, *Economic Journal* 75, p 493–517.

Becker Gary – Kenneth Murphy 1988. “A theory of rational addiction”, *Journal of Political Economy* 96:4 , p 675-700.

Becker Gary – Michael Grossman – Kenneth Murphy 1994. “An empirical analysis of cigarette addiction”. *American Economic Review*; 84(3):396-418.

Behrman, JR – A.B. Deolalikar 1988. ”Health and Nutrition”, in Chenery, Hollis – T.N. Srinivasan (red) *Handbook of development economics*. Vol. 1 New York ; Amsterdam :North-Holland

Cameron, Colin – Pravin Trivedi 2005. *Microeconometics. Methods and Applications*, Cambridge: Cambridge: University Press.
Chaloupka, Frank 1991. “Rational addictive behavior and cigarette smoking”, *Journal of Political Economy* 99, p 722-742.
Chaloupka, Frank. – K.E. Warner. 2000. "The economics of smoking", in A. J. Culyer and J. P. Newhouse (Eds), *Handbook of Health Economics* Amsterdam: Elsevier.
Chaloupka, Frank – Henry Wechsler 1997. “Price, tobacco control policies and smoking among young adults”, *Journal of Health Economics* 16, p 359-373.
Chamberlain, G 1980. ”Analysis of covariance with qualitative data”, *Review of Economic Studies*, 47, p 225-238.
Chamberlain, G 1984. ”Panel Data”, in Griliches, Z. and Intriligator M.D. (eds) *Handbook of Econometrics* Vol 2, Amsterdam: Elsevier.
Chaney E.F – D.K. Roszell 1985 “Coping in opiate addicts maintained on methadone.” In S Shiffman & T.A. Wills (eds) *Coping and substance use,* (p267-293).Orlando, FL; Academic Press.
Chen, Likwang - Matthew Farrelly 2001. "The relationship between the presence of children and men´s cigarette consumption"; in Michael Grossman and Chee-Ruey Hsieh (eds) *The Economic Analysis of Substance Use.* Cheltenham: Edward Elgar Publishing
Clark, Andrew – Fabrice Etilé 2002. “Do health changes affect smoking?” *Journal of Health Economics* 21, p 533-562.
Clark, A. - F Etilé – F Postel-Vinay – C Senik – K. Van Der Straeten. (2005) "Heterogeneity in Reported Well-being: Evidence from Twelve European Countries”, *Economic Journal* 115, C118-C32
Cohen, S – G.M. Williamsson 1988. “Percieved stress in a probability sample of the United States”. In S Spacapan & S Oskamp (eds), *The social psychology of health* (p 31-67) Newbury Park, CA: Sage
Cragg, J.G. 1971. “Some statistical models for limited dependent variables with application to the demand for durable goods”, *Econometrica* 39, p 829-844.
Deb, Partha - Pravin. Trivedi 2002. "The structure of Demand for health care: latent class vs two part models", *Journal of Health Economics* 21: p 601-625.
DiCecca, Philip – Don Kenkel – Alan Mathios 2000. "Racial Difference in the Determinants of Smoking Onset", *Journal of Risk and Uncertainty*, 21:2/3, p 311-340.
DiCicca, Philip – Don Kenkel – Alan Mathios 2008. “Cigarettes taxes and the transition from youth to adult smoking”, *Journal of Health Economics* 27:4, p 904-917.
Doll, R 2000. “Fifty years of research on tobacco,. *Journal of Epidemiological Biostat* 5:6, p 321-329.
Dorsett R - Marsh A 1998. *The Health Trap: Poverty, Smoking and Lone Parenthood*. London: Policy Studies Institute.
Dow, Will - Edward Norton. 2003. “Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions”, *Health Services & Outcomes Research Methodology* 4, p 5−18.
Duan, N 1983. “Smearing estimate a nonparametric retransformation method”, *Journal Of Americas Statistical Association.* 78, p 605–610.
Emery, Sherry - Martha M. White, John P. Pierce 2001. “Does cigarette price influence adolescent experimentation?”, *Journal of Health Economics* 20, p 261–270.
Folkhälsoinstitutet 2007. *Minskat bruk av tobak – var står vi idag*, Stockholm: Statens Folkhälsoinstitut:
Foster, M - Andrew Jones 2001. "The role of tobacco taxes in starting and quitting smoking: Duration analysis of British data", *JRSS(A) Statistics in Society* 164, p 517-547.
Garcia, J. – Jose M. Labeaga 1996. “Alternative Approaches to Modelling Zero Expenditure: An Application to Spanish Demand for Tobacco”, *Oxford Bulletin of Economics and Statistics* 58, p 489- 506.
Gerdtham, Ulf-G –Magnus Johannesson –L Lundberg -D Isacson 1999. ”The demand for health: results from new measures of health capital”, *European Journal of Political Economy* 15, p 501–521.
Gerdtham, Ulf-G – Magnus Johannesson 2002. “Do life savings Regulations save Lives?”, *The Journal of Risk and Uncertainty*, 24:3, p 231–249.

Gilleskie, Donna – Koleman Strumpf 2005. "The Behavioral Dynamics of Youth Smoking", *The Journal of Human Resures* 40:4, p 823-866.
Graham, H. 1993. *When Life's a Drag: Women, Smoking and Disadvantage*. London: Department of Health.
Grossman, Michael, 1972. "On the concept of health capital and the demand for health", *Journal of Political Economy* 80, p 223-355.
Grossman, Michael – Frank Chaloupka 1998. "The demand for cocaine by young adults: a rational addiction approach", *Journal of Health Economics* 17:4, p 427-474.
Gruber, Jonathan – Botond Köszegi 2001. "Is Addiction "Rational"? Theory And Evidence", *The Quarterly Journal of Economics* 116:4, p 1261- 1302.
Hammar, Henrik – Fredrik Carlsson 2001. "Smokers Decisions to Quit Smoking", *Working Papers in Economics* 59; Göteborg University
Heckman, James 1979. "Sample selection Bias as a Specification Error", *Econometrica* 47, p 153-161
Janzon, Ellis - Gunnar Engström, Martin Lindström, Göran Berglund, Bo Hedblad and Lars Janzon 2005. "Who are the ``quitters"? a cross-sectional study of circumstances associated with women giving up smoking*", Scandinavian Journal of Public Health* 33, p 175-182
Jones, Andrew 1989. "A double-hurdle model of cigarette consumption", Journal of Applied Econometrics 4:1, p 23-39.
Jones, Andrew 1994. "Health, addiction, social interaction and the decision to quit smoking". *Journal of Health Economics* 13, p 93-110.
Jones, Andrew 2000. "Health Econometrics". In Culyer, A. J. and Newhouse, J. P. (eds), *Handbook of Health Economics*. Amsterdam: Elsevier
Jones, Andrew 2006. *Applied Health Econometrics for health economists: A practical Guide 2nd ed*, Norfolk: Office of Health Economics
Jones, Andrew – Jose M. Labeaga 2003. "Individual heterogeneity and censoring in panel data estimates of tobacco expenditure". *Journal of Applied Econometrics* 18, p 157-77.
Jones, Andrew – Xander Koolman – Nigel Rice 2006. "Health related non-response in the BHPS and ECHP: using inverse probability weighted estimators in non-linear models", *Journal of Royal Statistical Society Series A,* 169: 543-569
Jones, Andrew – Nigel Rice – Teresa Bago d'Uva – Silvia Balia 2007. *Applied Health Economics*, Abingdon: Routledge,
Kamrul, Islam - Ulf-G Gerdtham - Philip Clarke – Kristina Burström 2007. "Does the concentration index of health change with aging of population" Evidence from Swedish panel data" Fourtcoming in Health Economics
Keele, L. J. – D.K. Park 2004 "Difficult Choices: An Evaluation ofHeterogenous Choice Models" *Paper presented at the annual meeting of the The Midwest Political Science Association, Palmer House Hilton, Chicago, Illinois* Online <.PDF> Retrieved 2008-06-28 from http://www.allacademic.com/meta/p83095_index.html
Kennedy, Peter 2003. *A Guide to Econometrics 5th ed*, Molden: Blackwell Publishing;
Labeaga, Jose .M. 1993. "Individual behaviour and tobacco consumption: a panel data approach", *Health Economics* 2, p 103-112.
Labeaga, Jose .M. 1999. "A Double Hurdle Rational Addiction Model with Heterogeneity: Estimating the Demand for Tobacco ". *Journal of Econometrics* 35, p 45-62
Laux, F.L. 2000. "Addiction as a market failure: using rational addiction results to justify tobacco regulation", *Journal of Health Economics*, 19, p 421-37.
Lindström, Martin 2004. "Psychosocial work conditions, social capital, and daily smoking: a population based study" *Tobacco Control* 13, p 289-295
Lindström, Martin 2007. "Nicotine replacement therapy, professional therapy, snuff use and tobacco smoking: a study of smoking cessation strategies in southern Sweden", *Tobacco Control* 16; p 410-416.

Lindström, Martin – Bertil Hanson Per-Olof Östergren – Göran Berglund 2000. “Socioeconomic differences in smoking cessation: the role of social participation”, *Scandinavian Journal of Public Health* 28:3, p 200 – 208.

Lindström, Martin – Sven-Olof Isacsson - S Elmståhl 2003b. “Impact of different aspects of social participation and social capital on smoking cessation among daily smokers: a longitudinal study” *Tobacco Control*,12, p 274-281.

Lindström, Martin - Sven-Olof Isacsson - The Malmö Shoulder-Neck Study Group 2002. "Smoking cessation among daily smokers, aged 45-69 years: a longitudinal study in Malmö, Sweden"; *Addiction* 97, 205-215

Lindström, Martin – Ellis Janzon 2007. “Social capital, institutional (vertical) trust and smoking: A study of daily smoking and smoking cessation among ever smokers”, *Scandinavian Journal of Public Health* 35; 460-467.

Lindström, Martin - Mahnaz Moghaddassi - Kristian Bolin - Björn Lindgren - Juan Merlo 2003a. “Social participation, social capital and daily tobacco smoking: a population-based multilevel analysis in Malmö, Sweden” *Scandinavian Journal of Public Health* 31, p 444-450

Lindström, Martin. - Jan Sundquist 2002. “Ethnic differences in daily smoking in Malmo, Sweden. Varying influence of psychosocial and economic factors”. *European Journal of Public Health* 12, 287-294.

Lindström; Martin - Per-Olof Östergren 2001. “Intermittent and daily smokers: two different socioeconomic patterns, and diverging influence of social participation” *Tobacco Control* 10; 258-266

López Nicolás, Angel 2001. "How Important are Tobacco Prices in the Propensity to Start and Quit Smoking? An Analysis of Smoking Histories from the Spanish National Health Survey," E*conomics Working Papers 548*, Department of Economics and Business, Universitat Pompeu Fabra

Lundborg, Petter 2006. “Having the wrong friends? Peer effects in adolescent substance use”, *Journal of Health Economics*; 25, p 214-233

Lundborg, Petter - Björn Lindgren 2004. “Do they know what they are doing? Risk perceptions and smoking behaviour among Swedish teenagers”. *Journal of Risk and Uncertainty* 28, p261-286.

Madden, David 2008. “Sample Selection Versus Two-Part Models Revesited: the Case of Female Smoking and Drinking”, *Journal of Health Economics* 27, p 300-307.

Manning, Willard 1998. “The logged dependent variable, heteroskedasticity, and the retransformation problem”, *Journal of health Economics* 17, p 283-295.

Manning, Willard – John Mullahy 2001. “Estimating log models: to transform or not to transform” *Journal of health Economics* 20, p 461-494

Mullahy, John 1998. “Much ado about two: reconsidering retransformation and the two part model in health econometrics” *Journal of health Economics* 17, p 247-281

Mundlak, Yari 1978. ”On the Pooling of Time Series and Cross Section Data,” *Econometrica* 46, 69-85

Muurinen, Jaana Marja 1982. “Demand for health: A generalised Grossman Model”, *Journal of Health Economics* 1, p 5-28

Nicoletti, C - F Peracchi 2005. ”A cross-country comparison of survey nonparticipation in the ECHP”, *Journal of Royal Statistical Society Series A* 168, p173-781

Nystedt, Paul 2006 "Marital life course events and smoking behaviour in Sweden 1980-2000" *Social Science & Medicine* 62: p 1427-1442

O’Donnell, Owen – Eddy van Doorslaer _ Adam Wagstaff _ Magnus Lindelöw 2008. *Analyzing health equity using household survey data.* The World Bank: Washington

Orphanides, A – D Zervos 1995. "Rational addiction with learning and regret", *Journal of Political Economy*, 103, 739-58.

Orphanides, A – D Zervos 1998. "Myopia and Addictive behaviour", *The Economic* Journal 108, 75-91

Pacula, R. L. 1997. “Economic modelling of the gateway effect *Health Economics*, 6, 521–4.

Pagan, A. R. – D. Hall 1983. “Diagnostic tests as residual analysis”, *Econometric Reviews* 2:159-218.

Pesaran, M. H – L. W. Taylor 1999. “Diagnostics for IV regressions”, *Oxford Bulletin of Economics and Statistics* 61: 255-281.
Pollak, RA 1970. “Habit formation and dynamic demand functions”, *Journal of Political Economy*; 78, p 745-763.
Pollak, RA 1976. “Habit formation and long-run utility functions”, *Journal of Economic Theory* 13, p 272-297.
Pollak, RA 1978. “Endogenous tastes in demand and welfare analysis”, A*merican Economic Review* 68, p 374-379.
Powell, Lisa - John A. Tauras - Hana Ross 2005. “The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior”, *Journal of Health Economic*s 24, p 950–968.
Raptou, Eleni – Konstadinos Mattas - Efthimia Tsakiridou – Constantinos Katrakilidis 2005. “Factors Effecting Cigarette Demand”, *International Advances in economic Research* 11, p 275-290
Rosenzweig, Mark – Paul Schultz 1983. “Estimating a Household Production Fucntion: Heterogeneity, the Demand for Health Inputs and their Effects on Birth Weights” *Journal of Political Economy* 91:5
Ross, Hana - Frank Chaloupka 2001. “The Effect of Cigarette Prices on Youth Smoking” *ImpacTeen Research Paper Series*, No. 7
SCB 1982. *Socioekonomisk Indelning; Meddelanden i samordningsfrågor(MIS)1982:4;* Statistics Sweden
SCB 2001. *Analys av bortfallet bland ‘‘utsatta’’ grupper i Underso¨kningen av levnadsförhållanden (ULF) 1986–1999*. (Analysis of falling off among ‘‘vulnerable’’ groups in the survey on living conditions (ULF) (1986–1999). Statistics Sweden. Stockholm.
SCB 2007 “Alkohol och tobaksbruk” *Levnadsförhållande Rapport 114*, Statistic Sweden
Schelling, TC 1984 “Self-command in practice, in policy, and in a theory of rational choice”. *American Economic Review* 1984;74:1-11.
Seshamani, Meena – Alastair M Gray 2004. "A longitudinal study of the effects of age and time to death on hospital costs," *Journal of Health Economics*, 23:2, pages 217-235,
Siahpush, M - JB Carlin 2006 “Financial stress, smoking cessation and relapse: results from a prospective study of an Australian national sample”. *Addiction;* 110:121–7.
Simon, H. 1978. “Rationality as process and as product of thought”. *American Economic Review* 68, 1-16.
Smith, Ian 2002. “European divorce laws, divorce rates, and their consequences” In A.W. Dnes, & R. Rowthorn (eds) *The law and economics of marriage and divorce* Cambridge: Cambridge University Press
Statacorp 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP
Stehr Mark 2007. “The Effect Of Cigarette Taxes On Smoking Among Men And Women” *Health Economics* 16, p 1333–1343
Stiegler, George – Gary Becker 1977. “De gustibus non est disputandum”, *American Economic Review* 67, p 76-90.
Suranovic, S. M. – R.S. Goldfarb – T.C. Leonard 1999. “An economic theory of cigarette addiction”, *Journal of Health Economics*, 18, p 1–29.
Surgeon General 1989. *Reducing the health consequences of smoking: 25 years progress - a report of the Surgeon General*. Rockville: U.S. Department of Health and Human Sevcices, Office on Smokin and Health
Tauras, John 2005. “An Emperical Analysis of adult cigarette demand” *Eastern Economic Journal* 31:3
Tauras, John – Frank Chaloupka 1999. “Determinants of smoking cessation: An analysis of young adult men and women” *Working Paper 7262 NBER Working Paper Series*
Thaler R - HM Shefrin 1981. “An economic theory of self control”. *Journal of Political Economy*;89 p 392-406.
Tooze, AJ – GK Grunwald – R Jones 2002. “Analysis of repeated measures data with clumping at zero”. *Statistical Methods in Medical Research*, 11, p 341-355
Wagstaff, Adam., 1986. “The demand for health: Some new empirical evidence”, *Journal of Health Economics* 5, p 195–233.
van Doorslaer, Eddy 1987. *Health, knowledge and the demand for medical care* Assen: Van Gorcum

Van Ourti, Tom 2004. "Measuring horizontal inequity in Belgian health care using a Gaussian random effects two part count data model," *Health Economics*, 13:7, p 705-724.

Wangen, Knut R. 2004. "Some Fundamental Problems in Becker, Grossman and Murphy's Implementation of Rational Addiction Theory" *Discussion Papers No. 375*, Statistics Norway, Research Department

Verbeek, Marno 2004. *A Guide to Modern Economterics 2nd ed*, Chichester: John Wiley & Sons, Ltd

Verbeek, Marno – T.E. Nijman 1992. "Testing for Selectivity Bias in Panel Data Models" *International Economic Review,* 33, p 681-703

Wills, Thomas Ashby 1990. "Stress and coping factors in the epidemiology of substance use". In L.T. Kozlowski, H.M. Annis, H.D. Cappell, F.B. Glaser, M:S: Goodstadt, Y Israel, H Kalant, E.M. Sellers , & E.R. Vingillis (eds), *Research advances in alcohol and drug problems* 10, New York: Plenum

Wills, Thomas Ashby - Elisabeth A. Hirky 1996. "Coping and Substance Use: A theoretical Model and Review of the Evidence" in Zeidner, Mosche and Enlder Nomran (eds) *Handbook of Coping* Nwew York: John Wiley & Sons, Inc

Winston, GC 1980. "Addiction and backsliding : a theory of compulsive consumption". *Journal of Economic Behavior and Organization* ;1:4, p 295-324.

Wooldridge, Jeffrey 2000. "*The initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity",* Mimeo, Michigan State University

Wooldridge, Jeffrey 2002. *Econometric Analysis of cross sectional and panel data*, Cambridge, MA :The MIT Press

Wooldridge, Jeffrey 2005. "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity" *Journal of Applied Eonometrics* 20 p 39-54

Yen, Steven T. – Andrew M. Jones 1996. "Individual cigarette consumption and addiction: a flexible limited dependent variable approach", *Health Economics* 5, p 105-117.

Zweifel, Peter – Friedrich. Breyer 1997. *Health Economics*. Oxford: Oxford University Press.

# Appendix A

## Expectations, marginal and incremental effects

Because of simplicity, the expressions below will be in general formulas without any indexing on *i* or *t*. The expressions handle the endogenous, exogenous and the Mundlak type variables analogously. Indexing on the coefficients, $\beta_m$, refers to the different equations. $\beta_1$ =participation equation; $\beta_{2G}$ =second equation using GLM, $\beta_{2R}$ =second equation using the retransformation in Baser (2007). The expected value of the two part model is in a general formula expressed as;

$$E(y \mid x) = \Pr(y > 0 \mid x) * E(y \mid x, y > 0)$$

and the marginal effect of a continuous variable, included in both parts, can be derived by using the product rule of derivatives;

$$\frac{\partial E(y \mid x)}{\partial x} = \Pr(y > 0 \mid x) * \frac{\partial E(y \mid x, y > 0)}{\partial x} + E(y \mid x, y > 0) * \frac{\partial \Pr(y > 0 \mid x)}{\partial x}$$

(compare Dow & Norton 2003)

For a continuous variables only included in the first part;

$$\frac{\partial E(y \mid x)}{\partial x} = \frac{\partial \Pr(y > 0 \mid x)}{\partial x} * E(y \mid x, y > 0)$$

$$\frac{\partial E(y \mid x)}{\partial x} = \Pr(y > 0 \mid x) * \frac{\partial E(y \mid x, y > 0)}{\partial x}$$

The incremental effect of the dummy variables can be defined as

$$E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0)$$ [26]

$$= \langle \Pr(y > 0 \mid x, x_d = 1) - \Pr(y > 0 \mid x, x_d = 0) \rangle * E(y \mid x, x_d = 0)..$$

$$.. - \Pr(y > 0 \mid x, x_d = 0) * \langle E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0) \rangle$$

The formulas need to be substituted depending on the models applied in the first and the second step. If applying the Probit as the binary model in the first step gives;

$$\Pr(y > 0 \mid x) = \Phi(\beta_1 x)$$

where $\Phi(\cdot)$ is the standard normal CDF, and $\phi(\cdot)$ will further on be notation for the standard normal PDF.

[26] Analogously, the incremental effects in the two steps are defined as $\Pr(y > 0 \mid x, x_d = 1) - \Pr(y > 0 \mid x, x_d = 0)$ and $E(y \mid x, x_d = 1) - E(y \mid x, x_d = 0)$ respectively.

The marginal effects in the Probit of a continuous variable can be computed as,

$$\frac{\partial \Pr(y > 0 \mid x)}{\partial x} = \Phi(\widehat{\beta}_1 x) * \widehat{\beta}_1$$

The second part, the linear regression conditional on positive outcomes, differ between the two approaches discussed in Chapter 5; GLM of y with a log link and a linear regression of ln(y). In the GLM setting, the expected value in logs conditional on positive outcomes is expressed as;

$$\ln(E(y \mid x, y > 0)) = \widehat{\beta}_{2G} x$$

and can easily be transformed to levels;

$$E(y \mid x, y > 0) = \exp(\widehat{\beta}_{2G} x)$$

(compare Manning & Mullahy 2001)

By simple derivation it follows that the marginal effect can be computed as;

$$\frac{\partial E(y \mid x, y > 0)}{\partial x} = \widehat{\beta}_{2G} * \exp(\widehat{\beta}_{2G} x)$$

(compare Basu & Rathouz 2005)

In the other approach, the retransformation problem is more complicated, especially since unknown heteroskedasticity is present (compare Duan 1983, Manning & Mullahy 2001; Norton & Ai 2000). Following Baser (2007) a heteroskedasticity robust smearing factor is used to transform the log of y into levels. The expected value in levels can be generalized as;

$$E(y \mid x, y > 0) = \exp(\beta_{2R} x + 0.5 * \sigma^2 h_i(x))$$

where $\sigma^2 h_i(x)$ is the variance ($Var(u \mid x)$) under heteroskedasticity. For an unbiased retransformation an estimate of $\sigma^2 h_i(x)$ is needed. Although other functions are possible, letting h(x) be exponential assures the variance to be non-negative. Practically, this is performed by regressing the log of the fitted residuals on the independent variables, x[27]. The exponential of the fitted value, $\exp(\widehat{\gamma} x)$, is then used as an estimate of $\sigma_{it}^2 h_i(x)$. The expected value in levels conditional on positive observations can then be written as;

$$E(y \mid x, y > 0) = \exp(\widehat{\beta}_{2R} x + 0.5 * \exp(\widehat{\gamma} x))$$

and derivation gives the marginal effect for a continuous variable;

[27] In the dynamic model where IV is employed, is the endogenous lagged dependent used.

$$\frac{\partial E(y \mid x, y > 0)}{\partial x} = \exp(\hat{\beta}_{2R} x + 0.5 * \exp(\hat{\gamma} x)) * (\hat{\beta}_{2R} + 0.5 \hat{\gamma} * \exp(\hat{\gamma} x))$$

# Appendix B

## Stata output of the original panel data models

Below are the Stata outputs from the Random effects models. This illustrates an example of the models presented in coefficients, and to show the discussed value of ***rho***.

### *Static Models*

```
Random-effects probit regression                Number of obs      =     12703
Group variable: lpnr                            Number of groups   =      5168
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       2.5
                                                               max =         3
                                                Wald chi2(30)      =   1622.15
Log likelihood  = -3835.3495                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     childd2 |  -.0484712   .0672657    -0.72   0.471    -.1803096    .0833672
     childd3 |  -.2404638    .070975    -3.39   0.001    -.3795722   -.1013554
     childd4 |   -.408957   .1038603    -3.94   0.000    -.6125194   -.2053946
       educ2 |  -.0103669   .0617744    -0.17   0.867    -.1314425    .1107087
       educ3 |  -.2114505   .0905734    -2.33   0.020    -.3889711     -.03393
       educ4 |  -.4318976   .0873299    -4.95   0.000    -.6030611   -.2607341
         im2 |   .2953773   .0960099     3.08   0.002     .1072014    .4835531
        male |  -.7910171     .05926   -13.35   0.000    -.9071645   -.6748697
       alone |   .3931304   .0595781     6.60   0.000     .2763593    .5099014
        sei2 |  -.0996364   .0712905    -1.40   0.162    -.2393632    .0400905
        sei3 |  -.0128396   .0747695    -0.17   0.864    -.1593851    .1337059
        sei4 |  -.1014392   .0827967    -1.23   0.221    -.2637178    .0608393
        sei5 |   .1293915   .1008422     1.28   0.199    -.0682557    .3270386
        sei6 |  -.0441279   .0975685    -0.45   0.651    -.2353587    .1471029
        sei7 |  -.4202913   .2051783    -2.05   0.041    -.8224333   -.0181493
        sei8 |  -.3723622   .1534069    -2.43   0.015    -.6730341   -.0716903
financestr~s |    .317722   .0705492     4.50   0.000      .179448    .4559959
        age2 |   .0492699   .0748407     0.66   0.510    -.0974151    .1959549
        age3 |  -.1861613    .112609    -1.65   0.098    -.4068709    .0345483
        age4 |  -.5814356   .1588255    -3.66   0.000    -.8927278   -.2701434
        age5 |  -1.029927   .1934149    -5.32   0.000    -1.409014   -.6508412
        age6 |  -1.359959   .2357715    -5.77   0.000    -1.822063   -.8978554
       lninc |  -.0669262   .0540129    -1.24   0.215    -.1727896    .0389373
   smokepast |   .1244273   .0031677    39.28   0.000     .1182189    .1306358
       wave2 |  -.5592696   .0567671    -9.85   0.000     -.670531   -.4480081
       wave3 |  -1.477922   .0923433   -16.00   0.000    -1.658912   -1.296933
      mlninc |  -.0259144   .0867436    -0.30   0.765    -.1959287    .1440999
        mage |  -.0622419   .0051738   -12.03   0.000    -.0723823   -.0521014
    southern |  -.1318336   .0617883    -2.13   0.033    -.2529364   -.0107308
    northern |   .0470852    .078049     0.60   0.546    -.1058879    .2000584
       _cons |   2.504538   .8449476     2.96   0.003     .8484708    4.160604
-------------+----------------------------------------------------------------
    /lnsig2u |    .103771   .0854457                     -.0636996    .2712415
-------------+----------------------------------------------------------------
     sigma_u |   1.053255   .0449981                      .9686521    1.145247
         rho |   .5259195    .021304                      .4840805    .5673977
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   494.45 Prob >= chibar2 = 0.000
```

```
Random-effects GLS regression                   Number of obs      =      3277
Group variable: lpnr                            Number of groups   =      1789
R-sq:  within  = 0.0649                         Obs per group: min =         1
       between = 0.1199                                        avg =       1.8
      overall = 0.1129                                        max =         3
Random effects u_i ~ Gaussian                   Wald chi2(30)      =    333.80
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
```

| lny | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| childd2 | -.0241982 | .0253271 | -0.96 | 0.339 | -.0738384 | .0254421 |
| childd3 | .0075499 | .0277759 | 0.27 | 0.786 | -.0468898 | .0619897 |
| childd4 | .0312446 | .041546 | 0.75 | 0.452 | -.050184 | .1126732 |
| educ2 | -.0145183 | .0243818 | -0.60 | 0.552 | -.0623058 | .0332692 |
| educ3 | -.0137031 | .0386209 | -0.35 | 0.723 | -.0893986 | .0619925 |
| educ4 | -.1125113 | .0389484 | -2.89 | 0.004 | -.1888488 | -.0361738 |
| im2 | .1226428 | .0450819 | 2.72 | 0.007 | .034284 | .2110017 |
| male | .1086359 | .0282646 | 3.84 | 0.000 | .0532384 | .1640335 |
| alone | .0937329 | .0226484 | 4.14 | 0.000 | .0493428 | .1381231 |
| sei2 | -.0304233 | .0270401 | -1.13 | 0.261 | -.0834209 | .0225743 |
| sei3 | -.0176262 | .0294228 | -0.60 | 0.549 | -.0752938 | .0400414 |
| sei4 | .0196804 | .0340774 | 0.58 | 0.564 | -.0471102 | .086471 |
| sei5 | -.020848 | .0419977 | -0.50 | 0.620 | -.103162 | .0614659 |
| sei6 | .0640384 | .0406885 | 1.57 | 0.116 | -.0157097 | .1437865 |
| sei7 | -.4103788 | .1070914 | -3.83 | 0.000 | -.6202742 | -.2004835 |
| sei8 | -.0511705 | .0598587 | -0.85 | 0.393 | -.1684913 | .0661504 |
| financestr~s | .0755505 | .0251745 | 3.00 | 0.003 | .0262095 | .1248915 |
| age2 | .0462689 | .0282469 | 1.64 | 0.101 | -.0090939 | .1016318 |
| age3 | .0454564 | .042579 | 1.07 | 0.286 | -.0379969 | .1289097 |
| age4 | -.0485339 | .0590363 | -0.82 | 0.411 | -.1642431 | .0671752 |
| age5 | -.1434331 | .0721947 | -1.99 | 0.047 | -.2849321 | -.001934 |
| age6 | -.261445 | .0880708 | -2.97 | 0.003 | -.4340605 | -.0888294 |
| lninc | -.0017323 | .0227862 | -0.08 | 0.939 | -.0463925 | .0429278 |
| smokepast | .0120145 | .001381 | 8.70 | 0.000 | .0093078 | .0147212 |
| wave2 | -.0689108 | .0213939 | -3.22 | 0.001 | -.1108421 | -.0269795 |
| wave3 | -.2725544 | .0353448 | -7.71 | 0.000 | -.341829 | -.2032798 |
| mlninc | .0633451 | .0390765 | 1.62 | 0.105 | -.0132435 | .1399336 |
| mage | -.0119996 | .0021698 | -5.53 | 0.000 | -.0162524 | -.0077469 |
| southern | -.049133 | .0285912 | -1.72 | 0.086 | -.1051706 | .0069047 |
| northern | -.0511888 | .0374867 | -1.37 | 0.172 | -.1246614 | .0222839 |
| _cons | 2.037653 | .371568 | 5.48 | 0.000 | 1.309393 | 2.765913 |
| sigma_u | .49938725 | | | | | |
| sigma_e | .35248144 | | | | | |
| rho | .66747075 | (fraction of variance due to u_i) | | | | |

## *Dynamic Models*

```
Random-effects probit regression                Number of obs      =      7216
Group variable: lpnr                            Number of groups   =      3992
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       1.8
                                                               max =         2
                                                Wald chi2(30)      =    674.25
Log likelihood  = -2067.8084                    Prob > chi2        =    0.0000
```

| y1 | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| childd2 | -.1977291 | .1049782 | -1.88 | 0.060 | -.4034827 | .0080244 |
| childd3 | -.3527306 | .1148115 | -3.07 | 0.002 | -.5777569 | -.1277042 |
| childd4 | -.5623126 | .1632116 | -3.45 | 0.001 | -.8822015 | -.2424237 |
| educ2 | -.0269576 | .0883276 | -0.31 | 0.760 | -.2000764 | .1461613 |
| educ3 | -.2733943 | .1338839 | -2.04 | 0.041 | -.535802 | -.0109866 |
| educ4 | -.4728347 | .1328859 | -3.56 | 0.000 | -.7332862 | -.2123832 |
| im2 | .0169077 | .1379546 | 0.12 | 0.902 | -.2534784 | .2872938 |
| male | -.0928951 | .0763427 | -1.22 | 0.224 | -.2425241 | .0567338 |
| alone | .2208348 | .0853543 | 2.59 | 0.010 | .0535435 | .3881262 |
| sei2 | -.0184538 | .1060918 | -0.17 | 0.862 | -.22639 | .1894823 |
| sei3 | .1057003 | .1065529 | 0.99 | 0.321 | -.1031395 | .3145401 |
| sei4 | -.1504744 | .1208565 | -1.25 | 0.213 | -.3873487 | .0864 |

| | | | | | | |
|---|---|---|---|---|---|---|
| sei5 | .1658945 | .1430233 | 1.16 | 0.246 | -.1144262 | .4462151 |
| sei6 | -.0416689 | .1328945 | -0.31 | 0.754 | -.3021373 | .2187994 |
| sei7 | -.4430061 | .2606097 | -1.70 | 0.089 | -.9537916 | .0677795 |
| sei8 | .1034391 | .3468081 | 0.30 | 0.766 | -.5762922 | .7831705 |
| financestr~s | .3296871 | .1107436 | 2.98 | 0.003 | .1126337 | .5467405 |
| age2 | .2345311 | .1128234 | 2.08 | 0.038 | .0134012 | .4556609 |
| age3 | .2223545 | .1519124 | 1.46 | 0.143 | -.0753882 | .5200972 |
| age4 | .1659724 | .2063657 | 0.80 | 0.421 | -.238497 | .5704418 |
| age5 | -.0824791 | .2435627 | -0.34 | 0.735 | -.5598532 | .3948951 |
| age6 | -.0751971 | .2883576 | -0.26 | 0.794 | -.6403676 | .4899734 |
| lninc | -.1319166 | .0744858 | -1.77 | 0.077 | -.2779061 | .014073 |
| wave3 | -.5717699 | .0957693 | -5.97 | 0.000 | -.7594744 | -.3840654 |
| ydlag | .8510852 | .1642241 | 5.18 | 0.000 | .5292119 | 1.172958 |
| mlninc | -.1072711 | .1347557 | -0.80 | 0.426 | -.3713874 | .1568452 |
| mage | -.0217778 | .0075343 | -2.89 | 0.004 | -.0365448 | -.0070109 |
| inyd | 2.355597 | .3787454 | 6.22 | 0.000 | 1.61327 | 3.097924 |
| northern | .0110602 | .1090769 | 0.10 | 0.919 | -.2027267 | .2248471 |
| southern | -.0345546 | .0872281 | -0.40 | 0.692 | -.2055186 | .1364093 |
| _cons | 1.598243 | 1.266861 | 1.26 | 0.207 | -.8847598 | 4.081246 |
| /lnsig2u | .3071575 | .28993 | | | -.2610947 | .8754098 |
| sigma_u | 1.166 | .1690291 | | | .8776149 | 1.549148 |
| rho | .5761913 | .0707994 | | | .4350946 | .7058701 |

Likelihood-ratio test of rho=0: chibar2(01) = 28.87 Prob >= chibar2 = 0.000

| | | | |
|---|---|---|---|
| G2SLS random-effects IV regression | | Number of obs = | 1638 |
| Group variable: lpnr | | Number of groups = | 1097 |
| R-sq: | within = 0.0025 | Obs per group: min = | 1 |
| | between = 0.4009 | avg = | 1.5 |
| | overall = 0.3387 | max = | 2 |
| | | Wald chi2(29) = | 310.55 |
| corr(u_i, X) | = 0 (assumed) | Prob > chi2 = | 0.0000 |

| lny | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| smokelag | .0486786 | .0059339 | 8.20 | 0.000 | .0370484 | .0603088 |
| childd2 | -.033279 | .0396722 | -0.84 | 0.402 | -.1110351 | .0444771 |
| childd3 | -.0151777 | .0442445 | -0.34 | 0.732 | -.1018955 | .07154 |
| childd4 | -.0221952 | .064624 | -0.34 | 0.731 | -.1488558 | .1044655 |
| educ2 | -.0082554 | .0311094 | -0.27 | 0.791 | -.0692287 | .0527179 |
| educ3 | -.0576133 | .048817 | -1.18 | 0.238 | -.1532929 | .0380664 |
| educ4 | -.0664728 | .0483149 | -1.38 | 0.169 | -.1611683 | .0282226 |
| im2 | -.006141 | .0470119 | -0.13 | 0.896 | -.0982825 | .0860006 |
| male | .0573588 | .0282193 | 2.03 | 0.042 | .0020499 | .1126677 |
| alone | .0600005 | .0301225 | 1.99 | 0.046 | .0009615 | .1190395 |
| financestr~s | .0754854 | .0390535 | 1.93 | 0.053 | -.001058 | .1520287 |
| sei2 | -.0376231 | .038596 | -0.97 | 0.330 | -.11327 | .0380237 |
| sei3 | .0012351 | .0382305 | 0.03 | 0.974 | -.0736952 | .0761654 |
| sei4 | .0264949 | .0457813 | 0.58 | 0.563 | -.0632348 | .1162246 |
| sei5 | .0016793 | .0546935 | 0.03 | 0.976 | -.105518 | .1088767 |
| sei6 | .0593442 | .0512207 | 1.16 | 0.247 | -.0410466 | .1597349 |
| sei7 | -.5288911 | .1381704 | -3.83 | 0.000 | -.7997001 | -.2580821 |
| sei8 | -.0245485 | .1239528 | -0.20 | 0.843 | -.2674915 | .2183945 |
| age2 | .0461553 | .0487469 | 0.95 | 0.344 | -.0493868 | .1416973 |
| age3 | .0521499 | .0652858 | 0.80 | 0.424 | -.0758079 | .1801077 |
| age4 | .0161926 | .0887877 | 0.18 | 0.855 | -.1578282 | .1902133 |
| age5 | -.0489064 | .1069123 | -0.46 | 0.647 | -.2584508 | .1606379 |
| age6 | -.1087728 | .126531 | -0.86 | 0.390 | -.356769 | .1392234 |
| lninc | -.0413633 | .0335363 | -1.23 | 0.217 | -.1070931 | .0243666 |
| wave3 | -.1323596 | .0331174 | -4.00 | 0.000 | -.1972685 | -.0674507 |
| mlninc | .0980999 | .0565585 | 1.73 | 0.083 | -.0127526 | .2089525 |
| mage | -.0063046 | .0030239 | -2.08 | 0.037 | -.0122314 | -.0003778 |
| northern | -.0105589 | .0399236 | -0.26 | 0.791 | -.0888077 | .06769 |
| southern | .0067878 | .0312893 | 0.22 | 0.828 | -.0545381 | .0681138 |
| _cons | 1.471887 | .4643808 | 3.17 | 0.002 | .5617173 | 2.382056 |