NBER WORKING PAPER SERIES

THE CAUSAL EFFECTS OF YOUTH CIGARETTE ADDICTION AND EDUCATION

Rong Hai
James J. Heckman

Working Paper 30304
http://www.nber.org/papers/w30304

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2022

A previous version of this paper was circulated as "A Dynamic Model of Health, Addiction, Education, and Wealth." We thank Siddhartha Biswas and Kurtis Gilliat for excellent research assistance. We thank Jérôme Adda, Victor Aguirregabiria, Joseph Altonji, Gadi Barlevy, Stéphane Bonhomme, John Cawley, Indraneel Chakraborty, Flavio Cunha, Michael Darden, Mariacristina De Nardi, Armin Falk, Eric French, David Frisvold, Chao Fu, Titus Galama, George-Levi Gayle, Martin Gervais, Donna Gilleskie, Limor Golan, Luojia Hu, Juanna Joensen, Arie Kapteyn, Donald Kenkel, John Kennan, Michael Keane, Gizem Kosar, Lance Lochner, Arnaud Maurel, Costas Meghir, Daniel L. McFadden, Philip Monagan, Ronni Pavan, John Rust, Peter Savelyev, Andrew Shephard, Steven Stern, Petra Todd, Anne Villamil, Alessandra Voena, seminar participants at Cornell University, Johns Hopkins University, Washington University at St. Louis, University of Chicago, University of California Riverside, University of Iowa, University of Miami, University of Rochester, Federal Reserve Bank of Chicago, Annual Health Econometrics Workshop, briq Workshop: Structural Analysis of Inequality at University of Bonn, Cowles Summer Conference on Structural Empirical Microeconomic Models at Yale University, Lindau Meeting on Economic Sciences, World Congress of Econometric Society, Biennial Meeting of the American Society of Health Economists, North American Winter Meetings of the Econometric Society, and Southern Economic Association Annual Meeting for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Causal Effects of Youth Cigarette Addiction and Education
Rong Hai and James J. Heckman
NBER Working Paper No. 30304
July 2022
JEL No. H23,I12,I18,I24,I28

ABSTRACT

We develop and estimate a life-cycle model in a rational addiction framework where youth choose to smoke, attend school, work part-time, and consume while facing borrowing constraints. The model features multiple channels for studying the reciprocal causal effects of addiction and education. Variations in endowments and cigarette prices are sources of identification. We show that education causally reduces smoking. A counterfactual experiment finds that in absence of cigarettes, college attendance rises by three percentage points in the population. A practical alternative of 40% additional excise tax achieves similar results. Impacts vary substantially across persons of different cognitive and non-cognitive abilities.

Rong Hai
Department of Economics
University of Miami
523 F Jenkins
Coral Gables, FL 33146
rhai@bus.miami.edu

James J. Heckman
Center for the Economics of
Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and IZA
and also NBER
jjh@uchicago.edu

A data appendix is available at http://www.nber.org/data-appendix/w30304

This paper seeks to answer two intertwined questions. Does addiction to cigarettes reduce educational attainment? Does education impact cigarette addiction? Understanding the long term impacts of addictive substances on human capital can help policymakers determine their posture towards regulation and taxation of such substances. If education helps reduce addiction, then it is another important benefit of education not presently considered in education policy.

Prima facie, the relation between smoking and education is evident. The left panel of Figure 1 shows that the earlier a youth starts smoking, the less education he obtains. In the other direction, the right panel shows that the more a person is educated, the lower is the probability that he smokes.

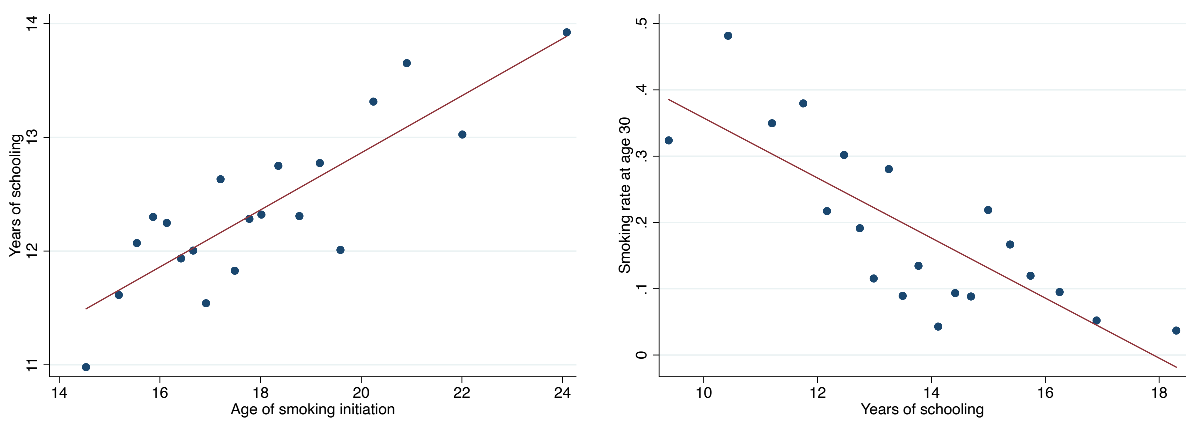


Figure 1: Relation between smoking and education

Note: Source: NLSY97 males. The left panel is a binscatter plot of the mean years of schooling as a function of age of smoking initiation among smokers. The right panel is a binscatter plot of the age-30 smoking rate as a function of years of schooling. Both plots control for cognitive and non-cognitive ability measures, and parental college education. Years of schooling is measured at age 30.

However, to establish causal effects of smoking on education, and of education on smoking, we face empirical challenges. First, a host of observable and unobservable characteristics can simultaneously determine education and cigarette smoking choices of young men (i.e., selection may be important). Second, once a person chooses education and/or addiction, their preferences regarding future investment in human capital evolves based on their past choices (i.e., dynamic selection). Thus, to estimate causal implications of education and addiction on each other over the lifetime, we need a life-cycle model of investment in education that embeds the choices regarding addiction.

Fortunately, two streams of literature come to our aid. Seminal work by Becker and Murphy (1994) provides a workhorse model of addiction decisions by individuals. The implications of the model have been empirically tested in the literature (Adda and Cornaglia, 2006; Arcidiacono et al., 2007; Darden, 2017). The model of rational addiction in which agents rationally choose to become addicted due to their preferences and continue to make optimal consumption decisions, is the first component for our framework. To these models, we add an analysis of their predictions for investments in education.

To bake in the second component of investment in education into a life-cycle model of addiction, we rely on the literature that starts with the seminal work of Keane and Wolpin (1997, 2001). Their model features education and savings decision in the presence of borrowing constraints and parental transfers over the life-cycle. Our model adjoins addiction decisions to their model and considers how smoking affects educational choices.

The inclusion of a model of the determinants of education serves to alleviate the first concern about selection—that education and smoking choices are jointly determined and also determined by other factors as well. Regarding the second concern of dynamic selection bias, we introduce endogenously determined time preferences that affect the co-evolution of education and addiction over the life-cycle of young men.

In our model, individuals start with differences in their early endowments (cognitive and non-cognitive abilities) as well as parental education. Parental education determines the transfers children receive to attend college as well as child preferences towards schooling. Cognitive and non-cognitive abilities also determine ex-ante preferences for schooling and smoking, as well as the initial discount factor. Forward-looking individuals start their lives and make dynamic decisions on smoking, schooling, normal goods consumption, savings, and borrowing over their life-cycles. The "addiction capital," accumulated based on agents' past smoking decisions, increases current smoking utility and makes smoking addictive (Becker and Murphy, 1988).

Three channels create a negative impact of smoking on education. First, addiction capital reduces an agent's discount factor. A lower discount factor reduces the incentives to invest in human capital. Second, our model allows for education to affect the flow utility of smoking. An agent anticipates the decline in utility of smoking due to enhanced education and thus, may reduce educational investment. Third, the monetary cost of smoking reduces resources available for tuition and fees.

Education also affects smoking decisions through similar channels. First, more education increases discount factors, which affect smoking decisions. Second, higher education reduces the flow utility of smoking. Third, borrowing constraints lead to substitution effects between schooling and smoking decisions. In addition, education increases earnings and hence has an income effect on smoking decisions.

Our model's primitives include cognitive ability, non-cognitive ability, parental education, and age. While we are focused on how these primitives affect education and smoking, these same primitives have been identified as determinants of friendships as well (see, Aboud and Mendelson, 1996, for a review).[1] For our purposes, peer effects are an intermediate channel through which the included primitives affect final outcomes of interest to us: addiction and education. Hence, our model does not separately include peer effects, as we already include the primitives.

We use data from the National Longitudinal Survey of Youth 1997 (NLSY97) over the time period 1997 to 2015 to estimate the model. Our data enable us to analyze the joint evolution of smoking addiction and education for each youth starting from a relative early age (age 15) over almost 20 years. It is well known that addiction to smoking starts early (Centers for Disease Control and Prevention (CDC), 2010). In addition to education and smoking decisions, we also observe each individual's employment, income, transfers received, parents' education, and various measures of cognitive and non-cognitive abilities. To estimate utility parameters, and thus the demand for addiction capital, we utilize state-year specific changes in cigarette prices and taxes. The resulting changes in prices facilitate changes in smoking behavior. The variation, thus created in smoking patterns, facilitates identification.

[1] Fletcher et al. (2014) show that parental education is a good predictor of friendship links. Burgess et al. (2011) also find that that academic achievement, personality, bad behavior, and mother's education are essential in the friendship formation process. Regarding selection on the primitive of gender: we use data from NLSY97 for males born between 1984 and 1982. Hence, our model is silent on gender.

We use a two-step procedure to estimate the model. In the first step, we estimate the parameters of the measurement system of cognitive and non-cognitive abilities using Maximum Likelihood Estimation (MLE). In the second step, we use the Simulated Method of Moments (SMM) to estimate parameters for individual preferences, earnings function, and discount factor. The initial conditions for cognitive and non-cognitive ability in the second step are obtained based on the parameter estimates in the first step. In total, we estimate 44 parameters in the second step, including parameters for the discount factor, preferences over smoking and schooling, and the earnings function. We have 105 targeted empirical moments that includes both life-cycle moments and coefficients from various regressions. Overall, our model fits the data well. For the 105 targeted moments, the average distance between the data moment and the corresponding model generated moment is 1.58 times the standard error of the corresponding data moment.

Our model is consistent with previously reported results.[2] We build on this literature in three ways. First, for each unit increase in "addiction capital," an individual at the mean level of cognitive and non-cognitive skills experiences a decrease in the discount factor by 0.038%. The effects of the subjective discount factor channel are cumulative over lifetime: as an example, for unit income and consumption increases in every period, the cumulative discounted utility of a smoker decreases by 0.95% (using the estimated baseline discount factor). For smokers, this lower expected utility generates lower investment in education as well, exacerbating outcome differences.

Second, our estimated consumption and addiction utility parameters produce the utility trade-offs facing smokers. A median 18-year old smoker, with an addiction stock of one unit receives more than three times additional utility from smoking a cigarette than from consumption of goods of the same dollar cost at median consumption levels.

Third, education reduces the utility of smoking. Four years of education counteracts the flow utility of smoking for the median smoker (with "addiction stock" of one unit). In this context of flow utility of smoking, the distinction between cognitive and non-cognitive skills is also very

[2]For example, we corroborate that individuals with high cognitive and non-cognitive abilities have higher discount factors and are more patient (Dohmen et al., 2011). Our discount factor estimate is also in line with the estimates in the literature (see, for example, Keane and Wolpin, 2001; Johnson, 2013; Blundell et al., 2016; Abbott et al., 2019, among others).

important. While cognitive skills have negligible effect on the utility from smoking, one standard deviation increase in non-cognitive skills has the same effect as one year of education in reducing the utility of smoking. In other words, individuals who fare better on dimensions such as self-control and discipline have less utility from smoking.

Using our estimated model, we can answer the intertwined questions that motivate our work. The first question we answer is: *does education causally reduce smoking?* To answer this question, we assign each individual one extra year of education at the formative age of 20. We find that this additional year of schooling reduces the smoking by four percentage points (21%) at age 20 (short-run effect). In the long-run by age 30, due to the additional year of education, the smoking rate is reduced by eight percentage points (31%) in the total population. The treatment effect is stronger among individuals with higher cognitive and non-cognitive skills. Ceteris paribus, those with higher non-cognitive skills are more likely to quit smoking in response to education.

The second question we answer is: *does cigarette addiction affect education?* To quantify the negative effect of smoking on education, we conduct a counterfactual analysis where smoking is eliminated as a choice. We find that if smoking is impossible, the college attendance rate increases by six percentage points among benchmark ever smokers—from 30 percentage points in the benchmark model to 36 percentage points. For the entire population, eliminating the possibility of smoking increases the college attendance rate by three percentage points, from 59 percentage points in the benchmark model to 62 percentage points in the counterfactual scenario. We find that a youth's non-cognitive ability is more instrumental in avoiding harmful behavior. However, once harmful behavior is not possible (in our counterfactual world), cognitive skills have a stronger effect on educational attainment. Policy impacts vary greatly by abilities.

Finally, we evaluate the impact of a more feasible policy on education: excise taxes that reduce addiction. Such taxation will not increase smokers' utility but can increase societal human capital investment. To preserve revenue neutrality, the excise tax revenue is redistributed as lump-sum per period tuition subsidy. We find that with an additional 40% excise tax, the college attendance rate reaches 36% among the would-be smokers in the benchmark model. This attendance rate is the same as the rate in the counterfactual above where we eliminated smoking as a choice. Under the 40% excise tax, the age-30 smoking rate is reduced to 24 percentage points, in comparison to 27 percentage points in the benchmark case. The implied elasticity of daily smoking participation is -0.28. We also find variations in policy impacts of the effects of excise taxes on smoking and education along dimensions of cognitive and non-cognitive abilities.

The addictive substance of choice has evolved over the last two decades. Compared to the NLSY97 youths who preferred cigarettes, today's youths are using marijuana and e-cigarettes. As long term economic studies on marijuana and e-cigarettes use are not yet possible, our results provide a window into the future of today's substance users as well.

This paper contributes to a growing literature that estimates structural life-cycle models of educational choices. Researchers have investigated the effects of parental transfers and credit constraints on youths' schooling choices (see, Keane and Wolpin, 2001; Johnson, 2013; Hai and Heckman, 2017; Abbott et al., 2019; Caucutt and Lochner, 2020). Building on these recent studies, we model the joint decision of schooling and smoking over a youth's life-cycle in the presence of parental transfers and borrowing constraints. We allow for selection into both smoking and education based on individual cognitive and non-cognitive abilities, parental education, and a youth's own endogenously accumulated wealth. Our model generates a causal impact of education on smoking that is consistent with empirical findings.[3] Our analysis complements existing studies by incorporating addiction as a choice, and showing that addiction in the formative years in the life of a youth has significant implications for his education decisions.

We also contribute to the rational addiction literature. A substantial body of research tests the empirical implications of the rational addiction model developed by Becker and Murphy (1988).[4]

[3]Grossman (1972), Kenkel (1991), Deaton (2003), Cutler and Lleras-Muney (2006), Grossman (2008), Currie (2009), Conti et al. (2010), Cutler and Lleras-Muney (2010), Goldman and Smith (2011), Deaton (2015), Heckman et al. (2018) and Savelyev and Tan (2019) find strong effects of education and personality skills on health and health behaviors.

[4]See, for example, Chaloupka (1991); Becker et al. (1991, 1994). Recent studies also estimate dynamic models of rational addiction (see, Gruber and Köszegi, 2001; Adda and Cornaglia, 2006; Arcidiacono et al., 2007; Adda and Lechene, 2013; Darden, 2017). Kenkel et al. (2002) build and calibrate a rational addiction model where peer influence is captured by the average addiction stock of an older generation. In their model, agents are homogeneous within each generation and do not make decisions on education.

The literature finds that forward-looking smokers adjust their smoking decisions in response to changes in cigarette prices and taxes, and that it is important to account for selection into smoking. Our model complements these studies in two aspects. First, we recognize that addiction in the early formative years of life has additional harmful effects on agents' education choices and human capital accumulation. Second, we estimate an endogenous discount factor that becomes smaller as a youth becomes more addicted. As a result, our model can account for the empirical findings that smokers are less patient and have less accumulated human capital and wealth.[5]

The structure of this paper is as follows. Section 1 presents our model. Section 2 discusses our data, model parameter identification, and estimation method. The parameter estimates and goodness of model fit are discussed in Section 3. We conduct counterfactual policy simulations to investigate the effects of education on smoking in Section 4.1, the impact of smoking on education in Section 4.2, and the impact of excise tax on smoking and schooling in Section 4.3. Section 5 concludes.

1 Model

We discuss the key ingredients of the model in Section 1.1. Section 1.2 sets up the individual utility maximization problem. Section 1.3 discusses the measurement equations used to proxy unobserved ability endowments.

1.1 Setup

1.1.1 Choice set

Agents are assumed to be economically active from age 15 to 64, i.e., $t \in [15, 64]$. In each period between age 15 and 64, individuals choose (i) consumption of a composite good $c_t > 0$ and next

[5] Orphanides and Zervos (1998) present a theoretical analysis of addiction where discount factor depends on addiction stock. They show such a model can account for addicts' disregard for the future consequences of their current actions. Kenkel and Wang (2001) theoretically examine the relationship between alcohol abuse and post-schooling human capital investment where the discount factor depends on addictive behavior. But both models are not estimated using data and there are no education decisions in their models.

period's stock of savings s_{t+1}; (ii) whether or not to smoke $d_{q,t} \in \{0,1\}$; (iii) whether to go to school $d_{e,t} \in \{0,1\}$; and (iv) whether to work part time while in school $d_{p,t} \in \{0,1\}$.[6] Agents can borrow up to their natural borrowing limits and hence the stock of savings in a period can be negative. Schooling choice ends before age 29. Agents derive utility or disutility over the aforementioned four choices.

The choice of age 15 as the starting period is based on data restrictions. Our NLSY97 sample starts with children at ages of 13, 14, and 15 in the initial interview year of 1997. For every year we go back past age 15, we will lose a third of our sample.[7] Starting age 65, agents retire and only make consumption and savings decisions till age 80. The relatively short horizon after retirement does not much affect our empirical estimates as our data are for the early portions of the life cycle when education decisions and smoking initiation begin.

1.1.2 Rational addiction

Following Pollak (1970); Becker and Murphy (1988); Orphanides and Zervos (1998), past cigarette consumption affects individuals' current utility of smoking through an "addiction stock." The law of motion of the addiction stock is as follows:

$$q_t = (1-\delta_q)q_{t-1} + d_{q,t-1}, \quad t \geq 1, \tag{1}$$

where the stock depreciates at rate $\delta_q \in [0,1]$, and the initial age addiction stock (at age 15) is zero, i.e., $q_{t_0} = 0$.

The presence of an addiction stock creates intertemporal non-separability in smoking utility and hence affects agents' future flow utility of smoking. We also allow addiction stock to affect how individuals discount future lifetime utility. As in the rational addiction literature, individuals are forward-looking and fully aware of the consequence of their current smoking decisions on

[6] Previous studies show that part time work is an important source of self-financing for school. See, for example, Keane and Wolpin (2001) and Johnson (2013).

[7] Note that we are only interested in daily smokers, not boys who tried smoking but did not become addicted. Among those who had ever smoked daily by age 35, only 5.5% began daily smoking before age 15 in our data.

future payoffs due to the working of the dynamics of the addiction stock.

1.1.3 Discount factor

Agent discount factors depend on addiction stock q_t (Becker and Mulligan, 1997; Orphanides and Zervos, 1998; Kenkel and Wang, 2001). This assumption is consistent with empirical evidence that addicts behave more myopically (Chaloupka, 1991). Addiction therefore affects education choices by reducing the perceived future value of the returns to education. The discount factor also depends on cognitive and non-cognitive abilities ($\boldsymbol{\theta} = \{\theta_c, \theta_n\}$) as shown in a recent work (Dohmen et al., 2011). We isolate the effect of addiction on discount factors separately from the effects arising from endowed abilities to evaluate the future. We postulate the following functional form for the discount factor $\beta(q,e,t,\boldsymbol{\theta})$:

$$\beta(q,e,t,\boldsymbol{\theta}) := \underbrace{\exp(-\phi_{e,t})}_{\text{survival rate}} \underbrace{\exp(-\kappa_q q)}_{\text{addiction effect}} \underbrace{\exp(-\kappa_r + \kappa_c\theta_c + \kappa_n\theta_n)}_{\text{endowed discount factor}}. \tag{2}$$

where $\exp(-\phi_{e,t})$ is the education and age-specific survival rate based on mortality data from CDC Vital Statistics and population data from American Community Survey.[8] In the data, actual survival rates increase with education and decrease with age. Incorporating $\phi_{e,t}$ into Equation (2) is a natural and parsimonious way to capture the notion that highly educated individuals are more patient and that individuals become less patient as they age.

The parameter $\kappa_q \geq 0$ captures the subjective belief of the harmful effects of addiction in increasing the probability of death. The term can also be interpreted as the difference in (subjective) life expectancy between addicts and non-addicts. As κ_q increases, addicted youths become more myopic and less patient. The parameter κ_r controls the level of subjective discount rates. κ_c and κ_n govern how the discount factor is shaped by agents' cognitive and non-cognitive abilities, respectively. All parameters in the discount factor are assumed to be known to the agent from the onset

[8] We calculate the mortality rate for each age and for five education categories: less than high school, high school, some college, 4-year college, and graduate degree. Appendix Figure A.3 reports the calculated survival probability for the five education categories at each age. Details of the calculation are reported in Appendix A.3.

of life cycle decision making.[9]

1.1.4 Flow Utility

Individuals have separable preferences over a composite good c_t, smoking $d_{q,t}$, schooling in a period $d_{e,t}$, and part-time work while in school $d_{p,t}$. Agents exhibit Constant Relative Risk Aversion (CRRA) utility with risk aversion parameter γ. An individual's preference function thus takes the following form:

$$\begin{aligned} U(c_t, \boldsymbol{d}_t; q_t, e_t, \boldsymbol{\theta}, \boldsymbol{X}_t, \boldsymbol{\epsilon}_t) &= \frac{{c_t}^{1-\gamma}-1}{1-\gamma} + \underbrace{u_q(d_{q,t}; q_t, e_t, \boldsymbol{\theta}, \varepsilon_{q,t})}_{\text{Smoking utility}} + \underbrace{u_e(d_{e,t}, d_{p,t}; \boldsymbol{X}_t, \boldsymbol{\theta}, \varepsilon_{e,t})}_{\text{Schooling utility}} \\ u_q(\cdot) &= d_{q,t} \cdot (\nu_0 + \nu_q q_t + \nu_e e_t + \boldsymbol{\nu}_\theta \boldsymbol{\theta} + \boldsymbol{\epsilon}_{q,t}) + \nu_{q^2} q_t^2 \\ u_e(\cdot) &= d_{e,t} \cdot (\xi_p d_{p,t} + \boldsymbol{\xi}_e \boldsymbol{X}_t + \boldsymbol{\xi}_\theta \boldsymbol{\theta} + \boldsymbol{\epsilon}_{e,t}), \end{aligned} \tag{3}$$

where vector $\boldsymbol{d}_t = (d_{q,t}, d_{e,t}, d_{p,t})$ represents the discrete choices regarding smoking, schooling, and part-time working while in school. Vector $\boldsymbol{\epsilon}_t = \{\varepsilon_{q,t}, \varepsilon_{e,t}\}$ represents preferences shocks associated with smoking and schooling choices.

Our specification of the flow utility of smoking $u_q(\cdot)$ follows Becker and Murphy (1988). As in their work, smoking utility initially increases with the addiction stock due to addictive preferences, but when the addiction stock becomes sufficiently large the smoking utility starts to decrease due to the harmful effects of smoking, i.e, $\nu_q > 0$ and $\nu_{q^2} < 0$. We extend the Becker-Murphy model by allowing years of schooling e_t and ability vector $\boldsymbol{\theta}$ to affect smoking utility. These effects could be due to knowledge of the harmful effects of smoking, self-control, and peer effects. We do not differentiate among the exact channels through which education, cognitive ability, and non-cognitive ability affect preferences for smoking.

The flow utility of education $u_e(.)$ captures agents' psychic costs and benefits of going to school. The vector $\boldsymbol{X}_t$ includes age, previous period's enrollment status $d_{e,t-1}$, indicator vari-

[9] In this paper, we assume a finite and certain upper limit to lifetimes. A better approach would be to assume an uncertain lifetimes. However, we do allow for uncertainty in lifetimes below age 80.

ables for a full-set of schooling levels and their interactions with age, and parental educational level e_{pr}. The flow utility of education $u_e(.)$ depends on the aforementioned vector $\boldsymbol{X}_t$, individual ability vector $\boldsymbol{\theta}$, and whether or not they work part time $d_{p,t}$ (Keane and Wolpin, 2001).

1.2 Model Solution

1.2.1 Individual's Optimization Problem

An agent's action space includes smoking, education, part-time working, savings, and consumption decisions. From age 15 to age 29, agents decide on schooling and part-time work. Once done with schooling, agents always work. Hence, there is no work choice after leaving school in the model. From age 65, a retired agent lives till age 80, but only chooses to consume. Let Ω_t denote the relevant information set at time t, including individual ability endowments, savings, education, addiction stock, market prices, earnings, parents' education, transfers from government and parents, and all the realized shocks. Thus an individual's optimization problem from initial age $t_0 = 15$ to age $T = 64$ is as follows:

$$V_t(\Omega_t) = \max_{\{\boldsymbol{d}_t, c_t, s_{t+1}\}} \left\{ U(c_t, \boldsymbol{d}_t; q_t, e_t, \boldsymbol{\theta}, \boldsymbol{X}_t, \epsilon_t) + \beta(q_t, e_t, t, \boldsymbol{\theta}) \cdot \mathbb{E}_t \left[V_{t+1}(\Omega_{t+1}) \middle| \Omega_t \right] \right\}$$

After 64, it becomes:

$$V_{65}(\Omega_{65}) = \max_{\{c_t\}} \sum_{t=65}^{80} \left\{ \left(\prod_{j=65}^{t-1} \beta(q_{64}, e_{64}, j, \boldsymbol{\theta}) \right) \frac{c_t^{1-\gamma} - 1}{1-\gamma} \right\} \quad \text{(Retirement Problem)}$$

subject to the budget constraint, borrowing constraint, and the law of motion of addiction stock q (Eq. 1). The discount factor $\beta(\cdot)$ is the period-specific subjective discount factor (Eq. 2).

1.2.2 Budget Constraint and Borrowing Constraint

An individual's budget constraint and borrowing constraint are respectively:[10]

$$\begin{aligned} \underbrace{s_{t+1} - s_t \cdot (1 + r(s_t))}_{\text{Savings in period } t} &= \underbrace{Y_t + T_t + G_t}_{\text{Earnings, transfers and subsidies}} - \underbrace{(c_t + p_t \cdot d_{q,t} + m_{e_{t+1}} \cdot d_{e,t})}_{\text{Consumption, addiction and education costs}} \\ s_{t+1} &\geq \underline{S}_{t+1} \end{aligned} \tag{4}$$

where rate $r(s)$ for borrowing, i.e, r_b, is higher than rate r_l if savings are positive. $Y_t(\cdot)$ is the earnings function discussed shortly.

Transfers: T_t denotes transfers from parents and is estimated outside the model.[11] We set parental transfers to be zero after age 30. G_t is a means-tested consumption subsidy from the government. In other words, government transfers G_t bridge the gap between an individual's "total cash" (i.e., positive savings $s_t(1+r_l)$, income Y_t and parental transfers T_t) and a consumption floor $\underline{c}$, as in De Nardi et al. (2010). Starting at age 65, G is education-specific social security income. Both the consumption floor and social security income are obtained outside the model (see Appendix Table B.III).

Prices: $m_{e_{t+1}}$ is net tuition and fees cost associated with education level e_{t+1}. To attend college, composite goods consumption c_t must exceed room and board costs $n_{e_{t+1}}$. Both tuition and fees cost m and room and board costs n are externally calibrated (see Appendix Table B.III). Let p_t denote the state-year specific annual monetary cost of smoking that varies by an individual's residence state and year. Variations in prices help to identify the model's smoking preference parameters. Individuals take prices as given and form expectations about next period's price changes

[10]The net tuition costs depends on schooling level that the youth is attending. Both the individuals' earnings function and parental transfers function depend on the individuals' choices, endowments, and time-varying outcome variables.

[11]Parental transfers T_t includes financial transfers for college tuition and fees and a consumption subsidy in the form of shared housing and meals if the agent is under age 18 or is attending high school. Following Kaplan (2012), we set the value of consumption subsidy to be \$7800 per year for the NLSY97 cohort. We assume a log-linear parental financial transfer function that depends on a youth's schooling and working decisions, education level, age, ability, and parental education. The parental financial transfer function is estimated outside the model. The results are reported in Appendix Table B.IV. Our specification of parental transfers captures paternalism and is consistent with the findings of previous research (see, for example, Keane and Wolpin, 2001; Johnson, 2013; Hai and Heckman, 2017; Abbott et al., 2019).

when making their dynamic optimal decisions. We assume that individuals' expectation of price growth is as follows:

$$\ln p_{t+1} - \ln p_t = \Delta_p + \varepsilon_{p,t+1} \tag{5}$$

where $\varepsilon_{q,t+1}$ is an i.i.d. forecasting error with standard deviation σ_p and is realized at the end of period t. The parameters Δ_p and σ_p are estimated outside the model using cigarette prices across U.S. states from years 1997 to 2015. (We estimate that $\Delta_p = 0.042$ and $\sigma_p = 0.088$ using cigarette price data. See Appendix A.2 for details.)

Savings: At age 15, savings are zero. Borrowing limit $\underline{S}_t$ depends on the maximum government student loan credit for a student's education level e_{t+1} ($\overline{L}^g(e_{t+1}, d_{e,t})$) and the agent's natural borrowing limit in the private debt market ($\overline{L}^s_t(e_{t+1}, \boldsymbol{\theta})$). The latter is the present value of future earnings conditional on education and ability. Maximum government student loan credit $\overline{L}^g(.)$ is externally calibrated in Appendix Table B.III. Section B.1 in the Appendix derives the agent's natural borrowing limit $\overline{L}^s_t(.)$.

1.2.3 Earnings function

Let $Y(e_t, k_t, t, \boldsymbol{\theta}, \varepsilon_{y,t})$ denote the earnings function if an individual works full-time after leaving school: it depends on education e_t, experience k_t, age t, cognitive and non-cognitive abilities $\boldsymbol{\theta}$, and are subject to an earnings shock $\varepsilon_{y,t}$. If a person works part-time while in school, then the earnings are $Y_t = 1/2 \cdot \exp(\omega_p) Y(.)$, where the parameter ω_p measures the pay penalty associated with part-time jobs while in school. We use an augmented Mincer earnings function for full time earnings, which is as follows:

$$\begin{aligned}
\ln Y(e_t, k_t, t, \boldsymbol{\theta}, \varepsilon_{y,t}) = {} & \omega_0 + \omega_{k,1} k_t + \omega_{k,2} k_t^2 + \underbrace{\omega_{k,3} k_t \mathbf{1}(e_t \geq 16)}_{\text{Higher return to experience by education}} + \underbrace{\omega_c \theta_c + \omega_n \theta_n}_{\text{Reward to ability}} \\
& + \underbrace{(\omega_{e,1} \mathbf{1}(e_t \geq 12) + \omega_{e,2} \mathbf{1}(e_t < 12))(e_t - 12) + \omega_{e,3} \mathbf{1}(12 \leq e_t < 16) + (\omega_{e,4} + \omega_{e,5}(e_t - 16))\mathbf{1}(e_t \geq 16)}_{\text{Effect by education level}} \\
& + \underbrace{\omega_{a,1}(18 - t)\mathbf{1}(t < 18) + \omega_{a,2}(20 - t)\mathbf{1}(18 \leq t \leq 20)}_{\text{Age effect}} + \varepsilon_{y,t}, \qquad (6)
\end{aligned}$$

where $k_t \geq 0$ is the years of potential experience and $\varepsilon_{y,t} \geq 0$ is a lower bounded idiosyncratic earnings shock. Following Mincer (1974), we calculate the years of potential experience as age minus years of education minus 6. We set the potential experience to be age minus 18 for high school dropouts. If an agent is younger than 18, we set k_t to zero. For example, the potential experience for those with $e_t \leq 12$ is zero at age 18 and grows each year after age 18.

As agents in our model start at a relatively young age (age 15) and $k_t = 0$ if age is less than 18, we introduce an additional parameter $\omega_{a,1}$ to capture wage growth over time before age 18. The age parameters are also important because our agents make smoking decisions at this young age. Hence matching budget constraints at ages less than 18 years is important.

1.2.4 Stochastic Specifications

We make the following functional form assumptions to facilitate computation and interpretation.

Preference Shocks The preference shocks to smoking ($\varepsilon_{q,t}$) and schooling ($\varepsilon_{e,t}$) in Equation (3) are assumed to be independently distributed over time and with each other. We also assume that the preference shocks follow the following normal distributions:

$$\begin{aligned}
\varepsilon_t = \{\varepsilon_{q,t}, \varepsilon_{e,t}\} \qquad & \varepsilon_t \perp\!\!\!\perp \varepsilon_{t'} \text{ all } t \neq t' \\
\varepsilon_{q,t} \sim N(0, \sigma_q^2) \qquad \varepsilon_{e,t} \sim N(0, \sigma_e^2) \qquad & \varepsilon_{q,t} \perp\!\!\!\perp \varepsilon_{e,t}
\end{aligned}$$

The parameters σ_q and σ_e are estimated within the model.

Forecast Errors We assume that the forecast errors of cigarette price in Equation (5) are independently distributed over time and follow a normal distribution as follows:

$$\varepsilon_{p,t+1} \perp\!\!\!\perp \varepsilon_{p,t'+1} \;\; \text{all } t \neq t'$$
$$\varepsilon_{p,t+1} \sim N(0, \sigma_p^2)$$

As discussed earlier in Section 1.2.2, we estimate $\sigma_p = 0.088$ using using cigarette price data outside the model.

Earnings Shocks We assume that the earnings shock $\varepsilon_{y,t} = \tilde{\varepsilon}_{y,t} - \mathbb{E}(\tilde{\varepsilon}_{y,t})$ where $\tilde{\varepsilon}_{y,t} \geq \underline{\varepsilon}_y = 0$ is drawn from a gamma distribution $\Gamma(a_y, b_y)$ (Chatterjee et al., 2007; Hai and Heckman, 2017). The gamma distribution allows us to separately model the shape and the scale of the earnings shock distribution which are governed by the parameters a_y and b_y, respectively. Under our specification, $\mathbb{E}(\tilde{\varepsilon}_{y,t}) = a_y b_y$ and $Var(\tilde{\varepsilon}_{y,t}) = a_y b_y^2$. The parameters a_y and b_y are estimated within the model.

1.3 Model's Initial Condition and Ability Measurement

We complete the model by specifying the initial conditions and the measurement equations relating the unobserved ability endowments to observed measures. In particular, at age 15, agents differ in their observed state variables—including parental education e_{pr}, age-14 education e_{14} and schooling status $d_{e,14}$—and unobserved, by the econometrician, cognitive and non-cognitive endowments ($\boldsymbol{\theta}$).

We do not directly observe cognitive and non-cognitive abilities at initial age 15. Instead, we observe proxies (Heckman and Robb, 1985, 1986). As Section 2.1 discusses in detail, we have four dedicated continuous measures for cognitive ability and six dedicated discrete measures for non-cognitive ability in the initial survey year of 1997. To incorporate both continuous and discrete measurements, we use linear models for continuous measures, probit models for discrete measures that are indicator variables, and ordered probit models for discrete measures that take three discrete values.

In particular, let $Z_{c,j}$ be the j-th observed measure of cognitive ability for $j \in \{1,\ldots,J_c := 4\}$ and $Z_{n,j}$ be the j-th observed measure of non-cognitive ability for $j \in \{1,\ldots,J_n := 6\}$. We assume that the following relationships hold for each measurement $j \in \{1,\ldots,J_k\}$ and $k \in \{c,n\}$:

$$Z^*_{k,j} = \mu_{z,k,j} + \alpha_{z,k,j}\theta_k + \boldsymbol{x}_z'\boldsymbol{\beta}_{z,k,j} + \varepsilon_{z,k,j} \tag{7}$$

$$Z_{k,j} = \begin{cases} Z^*_{k,j} & \text{if } Z_{k,j} \text{ is continuous} \\ \mathbf{1}(Z^*_{k,j} > 0) & \text{if } Z_{k,j} \text{ is an indicator variable} \\ \mathbf{1}(Z^*_{k,j} > 0) + \mathbf{1}(Z^*_{k,j} > \text{cutoff}_{k,j}) & \text{if } Z_{k,j} \text{ takes values } \{0,1,2\} \end{cases} \tag{8}$$

where $\boldsymbol{x}_z$ is a vector of individual control variables observed in the initial survey year 1997, including parental education, the youth's age in 1997, and the youth's schooling level in 1997. The measurement errors are assumed to be independently distributed across the equations and overtime. We also assume that the measurement errors follow normal distributions: $\varepsilon_{z,k,j} \sim N(0, \sigma^2_{z,k,j})$ for $k \in \{c,n\}$ and $j \in \{1,\ldots,J_k\}$.

We assume that the unobserved endowments in cognitive and non-cognitive abilities are jointly normally distributed and independent of all shocks specified in the model $\boldsymbol{\epsilon}_t$ at a point in time and over time. These skill endowments are either inherited from the parents or formed as a result of parental investment, due to parents' knowledge, preferences, or financial resources. We allow the mean of initial endowments to differ based on parental education e_{pr}, i.e., $\mu_k(e_{pr}) = \mu_{k,0} + \mu_{k,1}e_{pr}$, for $k = c,n$. In particular, the ex-ante joint distribution of the unobserved cognitive and non-cognitive abilities is assumed to be:

$$\begin{pmatrix} \theta_c \\ \theta_n \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_c(e_{pr}) \\ \mu_n(e_{pr}) \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \sigma_{c,n} \\ \sigma_{c,n} & \sigma_n^2 \end{pmatrix} \right), \tag{9}$$

where the standard deviations of cognitive and non-cognitive abilities are σ_c and σ_n, respectively in the ex-ante joint distribution, and $\sigma_{c,n}$ measures the covariance between cognitive ability and non-cognitive ability.

1.4 Summarizing the Model

Our model allows multiple channels for educational attainment and addiction to affect each other. First, the subjective discount factor changes based on addiction or education level: addiction increases myopia, and education decreases it. Second, education reduces the flow utility of smoking, forcing rational agents to choose between education and smoking. Third, the monetary cost of smoking (i.e. cigarette costs) is not negligible, especially for low-income youth. Thus, budget constraints may reduce valuable financial resources towards schooling or vice-versa. Finally, higher education leads to higher income, which shifts the budget constraint.

2 Empirical Strategy: Data, Identification, and Estimation

We discuss our data in Section 2.1. Identification is discussed in Section 2.2. Estimation is discussed in Section 2.3. Appendix B.2 summarizes the parameters that are externally specified.

2.1 Data

Our main data source is the National Longitudinal Survey of Youth 1997 (NLSY97). We obtain data on average cost and excise tax per pack of cigarette from the Centers for Disease Control and Prevention (CDC). We merge the cigarette price data with NLSY97 data based on individual states of residence in each year.

2.1.1 National Longitudinal Survey of Youth 1997

The NLSY97 is a nationally representative sample of approximately 9,000 youth born during the years 1980 through 1984. Over the sample period 1997 to 2015, NLSY97 provides extensive information on the respondents' smoking, schooling, employment, earnings, and monetary transfers from parents and government. It also provides information on cognitive ability measures, measures of behaviors and social emotional functioning of adolescents (i.e. non-cognitive ability), and parental education.

Appropriate Dataset: NLSY97 enables us to analyze individual smoking choices, schooling decisions, and earnings from a relatively early age (age 15), year by year. A limitation of the dataset is that sampled individuals are in their mid-30s at the end of sample period and we do not observe their choices afterwards. However, the vast majority of the smoking initiation occurs before age 21 (Choi and Stommel, 2017). By age 30, many unhealthy behavior patterns already form (Steinberg, 2008) and health disparities emerge (Conti et al., 2010). By age 30, most educational choices are complete. The NLSY97 is a valuable source of information for analyzing the smoking initiation and continuation decisions, formation of addiction capital, human capital investment, and wealth in adolescence and young adult life.

Sample Selection: We focus on the core cross-sectional sample of NLSY97. We use unweighted data because oversampling does not occur in the core sample (Johnson, 2013). We restrict our sample to males, so that our estimates are not confounded by fertility choices. We restrict our sample to individuals who were born in the year 1982 or later, so that they were 15 years or younger in the initial interview year 1997. This sample restriction ensures that we have information on respondent choices starting at the age of 15 before the onset of daily smoking.[12] We also drop observations for which information on smoking, years of schooling, parental education, and ability measures are missing. Our final sample is an unbalanced sample that contains 1,605 individuals from age 15 up to age 33.

Key Variable Definitions: Following CDC website, we measure smoking as daily smoking (i.e. regular smoking) in the data.[13] Youths' education is measured as the years of schooling completed. Parental education is the average years of schooling completed by parents. Parental transfers include total money transfers received from parents in each year, including allowance, non-allowance

[12]Among those who had ever smoked daily by age 35, 94.5% began daily smoking at age 15 or later in our data.

[13]See the CDC website: `https://www.cdc.gov/nchs/nhis/tobacco/tobacco_glossary.htm`. An every day smoker (previously called a regular smoker) is an adult who smokes every day and has smoked at least 100 cigarettes in his or her lifetime.

Table 1: Summary Statistics of Key Variables

Panel A: Youth smoking by education (age 30)				
	High school or less	College and more	difference	*p*-value of difference
Smoking	0.30	0.10	-0.20	0.00
Years smoked	4.88	1.64	-3.24	0.00
Never smoked	0.39	0.70	0.31	0.00

Panel B: Age-30 education by youth smoking history				
	Smoked before age 20	Never smoked before age 20	difference	*p*-value of difference
Years of schooling (age 30)	12.13	14.34	2.21	0.00
College and more (age 30)	0.34	0.68	0.34	0.00

Panel C: Key Variables By Age (NLSY97)				
	Age 15	Age 20	Age 25	Age 30
Years smoked as of t-1	0.00	0.71	1.80	2.90
Never smoked as of t-1	1.00	0.71	0.63	0.58
School enrollment	0.97	0.35	0.09	0.00
Working part-time while in school	0.13	0.26	0.06	0.00
Years of schooling	8.46	12.21	13.36	13.60
Earnings after leaving school	$2,488	$17,244	$29,957	$40,570

income, college financial aid gift, and inheritance. Additional details regarding variables used can be found in Appendix A.1.

Education and Smoking: Figure 1 points towards a relation between education and smoking. Panel A of Table 1 also shows that compared to individuals with less education, at age 30, individuals with a college degree are less likely to smoke, have lower accumulated years of smoking, and are more likely to be a never smoker. In the opposite direction, Panel B shows that youths who started smoking before age 20 have less years of schooling and are less likely to have a college degree at age 30 compared to youths who did not initiate smoking before age 20.

Key Summary Statistics by Age: Panel C of Table 1 reports the statistics of key variables over age groups. The average total number of years smoked is 0.71 at age 20 and grows to be 1.8 and 2.9 at ages 25 and 30, respectively. The second row of the panel shows that 71 percent men never smoked regularly before age 20 and 58 percent never smoked regularly before age 30.

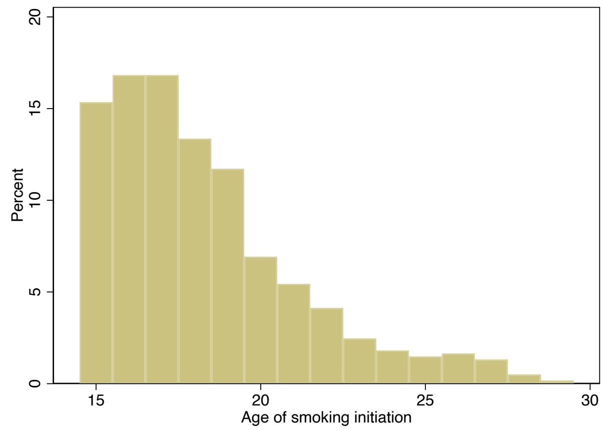


Figure 2: Age distribution of smoking initiation (among smokers)

Our paper underscores the impact of smoking in the formative years of youth on education. A majority of smokers initiated their regular smoking early in life. Figure 2 plots the age distribution of smoking initiation among youths who have smoked at or before age 30. About 15 percent of smokers initiated smoking at age 15. The smoking initiation rate is the highest at ages 16 and 17 and starts to decline afterwards. By age 21, 86 percent of smokers have already initiated their smoking.

The third row of Panel C in Table 1 reports that 97% of youth are enrolled in school at age 15. By age 20, the number drops to 35%. A bit more than a quarter of the 20 year olds who are in school work part-time. The average years of schooling (education) increases from 8.46 at age 15 to 13.6 at age 30. The average earnings after leaving school is $17,244 at age 20 and grows to be $40,570 at age 30. The earnings are in 2004 dollars.

Measures of Endowment: Appendix Table A.I reports measures of cognitive ability, measures of non-cognitive ability, parental education, and initial years of schooling. The measures of cognitive and non-cognitive abilities are used to extract factor scores of cognitive and non-cognitive skills using a factor model and measurement system (see Equations (7)–(9) in Section 1.3). Measures of cognitive ability are the four scores from the Armed Services Vocational Aptitude Battery (ASVAB), including arithmetic reasoning, mathematics knowledge, paragraph comprehension, and word knowledge. These cognitive ability measures have been standardized and have mean

0 and standard deviation one at age 15.

We use six measures for non-cognitive ability. The first two measures are two early age adverse behavior measures: (1) whether the youth had violent behavior towards others, and (2) whether the youth had stolen something worth at least $50. The early age behavior measures are valid measures for non-cognitive ability (see Heckman and Kautz, 2014 and Kautz and Zanoni, 2015 for detailed discussions). The remaining four measure are based on the youths' answers about how well do the following four statements describe themselves (on a three-point scale): you (1) "are unhappy, sad, or depressed," (2) "lie or cheat," (3) "don't get along with other kids," and (4) "have trouble concentrating or paying attention." These questions are from Achenbach's Youth Self Report (YSR; Achenbach, 1991) and help detect socio-emotional problems of adolescents. Achenbach and Ruffle (2000), Achenbach and Rescorla (2001), and Ebesutani et al. (2011) demonstrate the reliability and validity of YSR measures in assessing personality and behavioral problems and competencies of youths in both clinical and non-clinical settings. Individuals with high non-cognitive ability are less likely to display adverse behaviors and have less socio-emotional problems.

2.1.2 Cigarette Prices across States

We obtain average cost per pack of cigarettes and the federal and state tax per pack for each U.S. state and year between 1997 and 2015 from the CDC website (see AppendixFigure A.2).[14] In our sample, the median smoker smokes 14 cigarettes per day. Thus, the price of regular smoking for a youth p_q is the product of average cost per pack in the state-year (including federal and state taxes) times median consumption of 14 cigarettes times number of days in a year. Appendix A.2 provides additional details.

2.2 Identification

This section motivates sources of model identification. We first discuss identification of the parameters of the factor model and measurement system associated with the unobserved cognitive

[14]See "The Tax Burden on Tobacco, 1970–2018" by Orzechowski and Walker published on the CDC website: `https://chronicdata.cdc.gov/Policy/The-Tax-Burden-on-Tobacco-1970-2018/7nwe-3aj9`.

and non-cognitive abilities $\{\theta_c, \theta_n\}$. Then, we discuss the identification of smoking preference parameters and discount factor parameters. Finally, we discuss the identification of the remaining parameters of the model, including the earnings function, schooling preferences, and risk aversion.

2.2.1 Parameters on Measurement System of Abilities

The identification of factor models requires normalizations that set the location and scale of the factors (see Anderson and Rubin, 1956). Therefore, we normalize the unconditional means of factors $\{\theta_c, \theta_n\}$ to be zero, i.e., $\mathbb{E}_{e_{pr}}(\mu_c(e_{pr})) = \mathbb{E}_{e_{pr}}(\mu_n(e_{pr})) = 0$ in Equation (9) and set the scale of the factors by assuming $\alpha_{z,k,1} = 1$ for $k \in \{c, n\}$ in Equation (7). We further normalize the variances of the measurement error to one when we have dichotomous measurements. The factors are allowed to be correlated. Given these conditions, if there are three or more measures per factor, factor variances, covariances, and factor loadings are identified subject to the normalizations used to set the scale of each factor (see Williams, 2018).

Once the parameters of the factor model and measurement system are identified, we can obtain scores of unobserved cognitive ability and non-cognitive ability. These abilities will be used as covariates in the structural model identification and estimation below.

2.2.2 Smoking-specific Parameters

Identification of smoking-specific parameters requires a few relatively standard assumptions. First, all shocks $\boldsymbol{\epsilon}_t$ in the model are mutually independent and i.i.d over ages and $\boldsymbol{\theta}$. Second, utility is separable over smoking, schooling, and normal goods consumption (Eq. 3). Third, cigarette prices and excise taxes are exogenous to individuals' decisions and only affect individual decisions through budget constraints and future consumption decisions (Eq. 4). Fourth, individual smoking preferences do not depend on age directly once we condition on e_t, q_t, i.e., $u_q(\cdot)$ is a stable (time-invariant) function (an exclusion restriction). Under these assumptions, in our model, smoking-specific parameters are identified by exploiting: (1) variations in cigarette prices and taxes over time and across states and (2) variations in smoking rates over age.

Rational addiction creates non-time-separability in utility, i.e., current smoking decision depends on past and future smoking decisions. Chaloupka (1991) and Becker et al. (1994) show that if utility is non-time-separable, the demand for cigarette consumption in a period depends on prices in all periods through the effects of past and future prices on past and future consumption. Becker et al. (1994) derive a linear difference equation that provides a relation between current cigarette consumption with past and future smoking decisions along with the cost of cigarettes.

In our case, to aid the identification of smoking parameters, we utilize the following auxiliary model that relates present smoking decision $d_{q,t}$ (extensive margin), on past and future participation $(d_{q,t-1}, d_{q,t+1})$, cost of smoking p_t, education e_t, and abilities θ:

$$d_{q,t} = b_0 + b_p p_t + b_{q,-1} d_{q,t-1} + b_{q,+1} d_{q,t+1} + b_{q,e} e_t + b_{q,c} \theta_c + b_{q,n} \theta_n + \varepsilon_{q,t}. \tag{10}$$

The auxiliary model is estimated using two-stage least squares, where we instrument $(d_{q,t-1}, d_{q,t+1})$ with the lags and leads of excise taxes and cigarette prices following Chaloupka (1991) and Becker et al. (1994). The coefficients of this model are included as target moments in our model estimation, to aid identification.

Smoking Utility Preferences: The auxiliary regression model coefficient $b_{q,-1}$ captures the persistence between current smoking decision and past smoking decisions and helps us identify the addiction reinforcement parameter ν_q. The coefficients $b_{q,e}$ captures the variation in smoking rates by years of schooling and help to identify the smoking preference parameters ν_e. The coefficients $(b_{q,c}, b_{q,n})$ capture the covariance between smoking and abilities and help identify the smoking preference parameters (ν_c, ν_n).

Exploiting age as an exclusion restriction in smoking preferences (see Eq. 3).[15] We identify the preference shock parameter σ_q by matching the changes in smoking initiation rate over different ages. In particular, recall that the smoking initiation rate is more than 10 percent annually before

[15]Even though we assume that smoking preference function is a stable function that does not directly depend on age, age/maturation effects exist in our model and affect individuals' smoking decision in the following three channels. First, youth become "wiser" or more informed over time as their accumulation of schooling increases with age. Second, mortality risk increases as individuals grow older biologically. Last, the present discounted value of smoking changes over the life-cycle as the decision horizon shrinks with age.

age 20 and starts to decline sharply afterwards (see Figure 2). In our model, the magnitude of preference shocks σ_q affects the smoking initiation decisions. If idiosyncratic preference shocks are small, then choices are made very early based on endowment. If σ_q is very large, then smoking initiation rate is flatter over the life-cycle. Thus, the magnitude of σ_q helps determine the decline of initiation rates over the life-cycle. Furthermore, the average smoking initiation rate helps pin down the intercept parameter ν_0. Finally, the observed age pattern of year of smoked helps to identify the parameter ν_{q^2}.

Addiction Stock Depreciation: The coefficient b_p helps identify the addiction stock depreciation rate parameter δ_q. This is because, as Gruber and Köszegi (2001) show, the price elasticity of cigarette use (i.e., b_p) depends on the depreciation rate δ_q: When the depreciation rate is low, i.e., addiction stock q is low, then the price response is larger.

Subjective Discount Factor: The ratio of future consumption decision parameter to past consumption decision parameters: $b_{q,+1}/b_{q,-1}$ helps inform us about the smokers' subjective discount rate (conditional on right-hand-side variables in Equation 10), as discussed in Chaloupka (1991) and Becker et al. (1994). In our context, this ratio helps identify the effects of addiction on time preference parameter κ_q (Eq. 2).

We explore the variations in net worth distribution to pin down the remaining subjective discount factor parameters. The median net worth level helps to identify the intercept parameter of discounting function κ_r. Finally, the correlation between net worth and cognitive ability helps identify κ_c and the correlation between net worth and non-cognitive ability helps identify κ_n.

2.2.3 Parameters on Risk Aversion, Earnings, and Schooling Preferences

Our arguments for the identification of the parameters on risk aversion, earnings function, and schooling preferences are relatively standard.

Risk Aversion Parameter: In the presence of borrowing constraints and uninsured labor income risk, risk aversion parameter γ controls the size of the precautionary savings incentive. Thus, the risk aversion parameter γ is identified by the wealth inequality measure. We use the Gini coefficient of the net worth distribution as the measure.

Earnings Function: Under the assumption that the earnings shock $\varepsilon_{y,t}$ is independent of $\{\theta_c, \theta_n\}$ (Eq. 6), the structural parameters in the earnings function are identified by using the scores of unobserved factors $\{\theta_c, \theta_n\}$ as covariates (Heckman and Navarro, 2007; Heckman et al., 2018). The scores of unobserved factors are identified from the analysis of Section 2.2.1.

Schooling Preferences: Individuals with different "treatment times" (attained levels of schooling) have different outcomes while in school and after. The choice probability of schooling enrollment therefore identifies the schooling preference parameters. The identification argument here follows Heckman and Navarro (2007) and Navarro and Zhou (2017), for example.

2.3 Estimation Method

We use a two-step procedure to estimate the model. In the first step, we estimate the parameters in the initial conditions of the model, i.e., parameters on the factor model and measurement system (see Equations (7) to (9) in Section 1.3).[16] We maximize the following likelihood function:

$$\max \ \Pi_i \int_{\theta_c, \theta_n} f(Z_{i,c}, Z_{i,n}; \boldsymbol{x}_{i,z}, \theta_c, \theta_n) dF(\theta_c, \theta_n), \tag{11}$$

where $Z_{i,c} = \{Z_{i,c,1}, \ldots, Z_{i,c,J_c}\}$ is a vector of individual i's measures for cognitive ability, $Z_{i,n} = \{Z_{i,n,1}, \ldots, Z_{i,n,J_n}\}$ is a vector of individual i's measures for non-cognitive ability; $\boldsymbol{x}_{i,z}$ represents other controls in the measurement system (see Hai and Heckman, 2017, for details on this procedure).

In the second step, we use a Simulated Method of Moments (SMM) approach to estimate parameters for individual preferences, earnings functions and discount factors.[17] The initial conditions for cognitive and non-cognitive ability in the second step are obtained by simulation using

[16] The parameters that need to be estimated in this step include: (i) parameters of the initial distributions of cognitive ability, and non-cognitive ability: ($\mu_{c,1}$, $\mu_{n,1}$, σ_c, σ_n, $\sigma_{c,n}$), and (ii) parameters of measurement system: $(\mu_{z,c,j}, \alpha_{z,c,j}, \beta_{z,c,j}, \sigma_{z,c,j})_{j=1,\ldots,J_c}$, $(\mu_{z,n,j}, \alpha_{z,n,j}, \beta_{z,n,j}, \sigma_{z,n,j})_{j=1,\ldots,J_n}$.

[17] The choice variables in the model include not only discrete variables such as schooling and smoking but also continuous variables such as asset levels. As a result, we use Simulated Method of Moments (SMM) to estimate the model.

the parameters estimated in the first step.[18] In particular, we make 30 copies for each of the 1,605 individuals in NLSY97 data. In total, we simulate 48,150 individuals in the second step from age 15 to age 30, with a total of 770,400 observations in the simulated dataset.

Let $\hat{M}_N$ denote a vector of empirical moments calculated from a sample of data of size N and by $M_{\tilde{N}}$ a corresponding vector of moments predicted by the model. With $\tilde{N}$ simulations, the objective function of our SMM criterion is given by:

$$J := \min\ (\hat{M}_N - M_{\tilde{N}}(\text{parameters}))'\hat{W}_N(\hat{M}_N - M_{\tilde{N}}(\text{parameters})). \tag{12}$$

The weighting matrix is the inverse of the diagonal of the variance-covariance matrix of the data moments (Altonji and Segal, 1996).

In total, we estimate 44 parameters in the second step, including parameters for the discount factor, preferences over smoking and schooling, and the earnings function. We have 105 targeted empirical moments, $\hat{M}_N$, includes both life-cycle moments and coefficients from various regressions. Appendix Tables C.VIII to C.XI list all 105 targeted moments and the corresponding model fit.

3 Estimation Results and Goodness-of-Fit

This section reports the parameter estimates in Section 3.1. Section 3.2 reports the goodness-of-fit of the model.

3.1 Estimation Results

3.1.1 Initial Endowment of Skills

We start by reporting estimates of skill endowments. The parameter estimates of the factor model and the measurement system associated with skills $\{\theta_c, \theta_n\}$ are reported in Appendix Tables C.V

[18] Note that the parental education and age-15 years of schooling are directly observed in the data. We also assign net worth to be zero at age 15.

Table 2: Subjective Discount and Addiction Parameter Estimates

Note: Depreciation rate δ_q of addiction stock is from Eq. 1: $q_t = (1-\delta_q)q_{t-1} + d_{q,t-1}$. Subjective discount factor terms are presented in Eq. 2: $\beta(q,\boldsymbol{\theta}) := \exp(-\kappa_r + \kappa_c\theta_c + \kappa_n\theta_n)\exp(-\kappa_q q)$. Smoking preference terms $u_q(.)$ are presented in Eq. 3: $u_q(\cdot) = d_{q,t}\cdot(\nu_0 + \nu_1 q_t + \nu_e e_t + \boldsymbol{\nu}_\theta\boldsymbol{\theta} + \varepsilon_{q,t}) + \nu_2 q_t^2$.

Description	Symbol	Estimate	Standard Error
Depreciation Rate of Addiction Stock	δ_q	0.4330	0.0049
Subjective discount rate terms:			
Subjective discounting rate	κ_r	0.0249	0.0009
Myopia effect of addiction stock	κ_q	0.00038	0.000046
Effect of cognitive ability on subjective discount factor	κ_c	0.0097	0.0013
Effect of non-cognitive ability on subjective discount factor	κ_n	0.0113	0.0019
Addiction utility parameters:			
Addiction Stock	ν_q	0.0205	0.0014
Addiction Stock Squared	ν_{q^2}	-0.0022	0.0009
Yrs of Schooling -9	ν_e	-0.0048	0.0004
Cognitive Ability	ν_c	-0.00001	0.0016
Non-cognitive Ability	ν_n	-0.00816	0.0020
Intercept	ν_0	0.0030	0.0028
S.D. of Preference Shock to Smoking	σ_q	0.0238	0.0022

and C.VI. Table C.V reports the estimation results for Eq. 9. It shows that, on average, a youth whose parents have a college degree has higher non-cognitive ability. Furthermore, the correlation between youth's cognitive and non-cognitive abilities is positive, but moderate ($\sigma_{c,n}/(\sigma_c\sigma_n) = 0.39$). Heckman et al. (2016) report an estimate of 0.40 for a related model.

Using the parameter estimates from the first step, we obtain factor scores of cognitive ability and non-cognitive ability and use them as the initial distribution for the structural model simulation. Specifically, the standard deviation of cognitive ability factor is 0.74 and the standard deviation of the non-cognitive ability factor is 0.51.

3.1.2 Subjective Discount and Addiction Parameter Estimates

The main estimated parameters of interest are associated with cigarette addiction: The depreciation rate δ_q of addiction stock (Eq. 1), the subjective discount factor terms (Eq. 2), and coefficient vector ν in the smoking preference terms $u_q(.)$ (Eq. 3). Table 2 reports our parameter estimates.

Addiction Stock Depreciation: The estimated depreciation rate of cigarette addiction stock is $\delta_q = 0.43$, i.e. addiction stock depletes over a bit more than two years. This number for cigarettes is in line with medical observations regarding individual outcomes.[19] As mentioned before, the model can also be used to understand the effects of a more addictive substance, with a lower depreciation rate, for example.

Subjective Discount Factor: We estimate the subjective discount factor parameter κ_r to be 0.0249, which implies a subjective discount factor of 0.9754 among young non-addicts with average cognitive and non-cognitive abilities. Our estimates are in line with the estimated/calibrated discount factor values in previous structural models of educational choices. For example, Keane and Wolpin (2001) estimate the annual discount factor to be between 0.9758 and 0.9922. Johnson (2013) sets the annual discount factor to be 0.97. Abbott et al. (2019) estimate the annual discount factor to be 0.951. Blundell et al. (2016) sets the discount factor to be 0.98.

A key estimate is the subjective effect of each unit of addiction stock: $\kappa_q = 0.038\%$. The discount rate applied by a smoker compared to a non-smoker in any one period is not very different. However, effects are cumulative. A back of the envelope calculation shows that over the lifetime of an individual from age 15 to age 80, assuming unit income and consumption per period, the cumulative discounted utility decreases by 0.95%.[20] This reduction is not the full effect but only one of the channels that leads to differential outcomes conditional on smoking. For smokers, the lower expected utility feeds into lower investment in education as well.

Our parameter estimates suggest that individuals who have higher cognitive and non-cognitive abilities are more patient and care more about the future ($\kappa_c > 0, \kappa_n > 0$). These estimates are consistent with the findings in a recent work by Dohmen et al. (2011).

Smoking Preferences: The parameter ν_q captures the estimated utility benefits of continuing to smoke next year, conditional on having smoked in the past (Eq. 1). Table 2 shows that if the addic-

[19]Centers for Disease Control and Prevention (CDC) provides information about the timeline of how quitting smoking improves health outcomes. Around 1–2 years risk of coronary heart disease falls sharply. See `https://www.cdc.gov/tobacco/quit_smoking/how_to_quit/benefits/index.htm`.

[20]The present value of 1 unit of consumption over 65 years discounted at 0.038% in addition to the base discount rate κ_r is 99.05%.

tion stock q increases by one unit, then the additional utility from smoking next year is $v_q = 0.0205$. The squared term v_{q^2} has a value of -0.0022, which introduces weak concavity in the utility function from the addiction stock. Given these values of v_q and v_{q^2}, the accumulated addiction stock needs to be higher than 9 units in order for the disutility of addiction stock to dominate the positive effects of addiction stock from smoking. The weak concavity is reasonable because smoking does not lead to rapid harm.

To illustrate the magnitude of the parameter estimates associated with addiction stock in the utility function, consider the following example. At age 18, the median addiction stock among smokers is one. Compared to those that never smoke whose addiction stock is zero, one unit of addiction stock increases the flow utility of smoking by $v_q q - v_{q^2} q^2 = 0.0183$, for $q = 1$. This utility is equivalent to consuming 32% additional composite goods for age-18 smokers with median consumption.[21] Appendix A.2 reports data on annual cost of cigarette consumption. The average cost of cigarette consumption is \$1,042. Thus, we note that cigarette consumption is an example of rational addiction: dollar for dollar, the example 18-year old individual receives more than three times additional utility from cigarettes than from composite good consumption at the median consumption level. Our paper focuses on the *long term* costs of this *short term* trade-off in favor of cigarettes, as individuals reduce human capital investments due to rational addiction.

Table 2 shows that education and non-cognitive ability reduce the utility from smoking, while cognitive ability has an effect that is economically and statistically indistinguishable from zero ($v_c \approx 0$). Comparing the estimated magnitudes with v_q and with each other provides some additional insights. The addictive effects of one year of smoking (through $v_q q$, where $q = 1$) is offset by four additional years of education ($v_q / v_e \approx 4$).

Individuals who fare better on dimensions such as self-control and discipline have less utility from smoking. A one standard deviation increase in non-cognitive skills reduces the addictive effects of one year of smoking by about 20% ($\sigma_{\theta_n} v_n / v_q$). In comparison, as mentioned above, cognitive skills, separate from education, do not have a direct impact on smoking preferences. Our

[21]The median consumption of age-18 smokers is \$9,787 dollars. An additional 32% consumption creates the same utility from the composite good as $v_q q - v_{q^2} q^2$ for $q = 1$.

estimation results shed light on the importance of education and socio-emotional skills in reducing addiction.

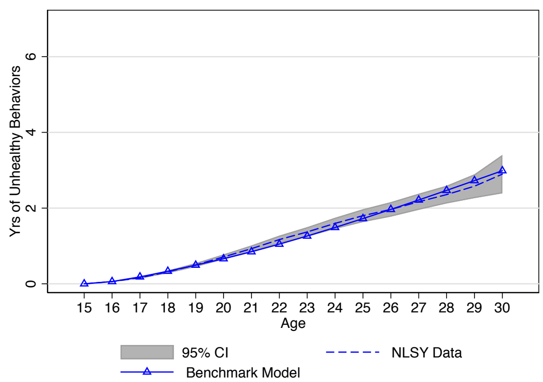
3.1.3 Additional Estimates

Appendix Table C.VII reports additional estimated parameters. Our estimate of relative risk aversion coefficient is 2.1290, which is within the range of estimates in the micro data studies on consumption and savings (Browning et al., 1999; Lochner and Monge-Naranjo, 2011).

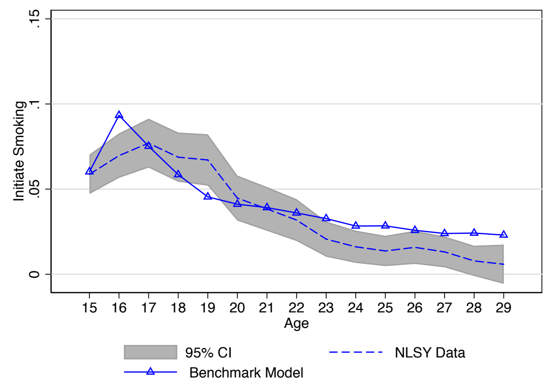
The estimated values of the schooling preference parameters are reported in Appendix Table C.VII Panel B ($u_e(.)$, Eq. 3). The psychic benefit of schooling is higher for individuals with higher cognitive and non-cognitive abilities. Note that cognitive ability is more than three times more important for schooling, i.e., $(\xi_{\theta,c}\sigma_{\theta_c})/(\xi_{\theta,n}\sigma_{\theta_n}) \geq 3$.

Individuals who have parents with college education also have higher flow utility of schooling $\xi_{e,pr}$. Parental college education has an effect that is more important than the effect of a two standard deviations increase in cognitive ability ($\xi_{e,pr}/(\xi_{\theta,c}\sigma_{\theta_c}) \geq 3$). The psychic costs of returning to school, $\xi_{e,r}$, for individuals who left school is positive and sizable. Working part-time while in school also impose a negative psychic cost while in school. Finally, the average psychic cost of schooling increases with the level of education, and is highest in graduate school, i.e., $\xi_{e,3}$. To compare magnitudes, the psychic cost of attending graduate school equals the utility of more than three standard deviations of cognitive ability $\xi_{e,3}/(\sigma_{\theta_c}\xi_{\theta,c}) \geq 3$.

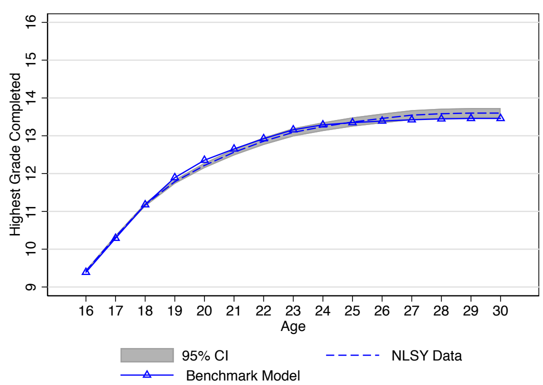
The parameter estimates for the earnings equation are reported in Appendix Table C.VII Panel C. Earnings increase with work experience but at a decreasing rate $\omega_{k,1} > 1, \omega_{k,2} < 0$. $\omega_{k,3}$ suggests that the return to experience is higher by 2.8% per year for individuals with a 4-year college degree. Our estimates show that if the cognitive ability is increased by one standard deviation (σ_{θ_c} = 0.74), earnings increase by 9.7%. Furthermore, a one standard deviation improvement in non-cognitive ability (σ_{θ_n} = 0.51) increases earnings by 4.1%. Thus, cognitive ability is more important for earnings.



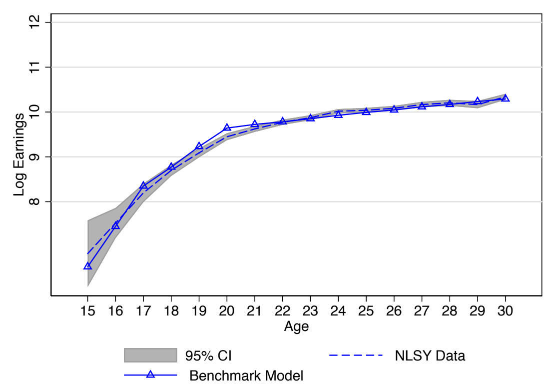
(a) Mean years of smoked over age



(b) Rate of smoking initiation over age



(c) Mean years of schooling over age



(d) Mean logarithm of earnings over age

Figure 3: Model Fit on Life-cycle Moments

3.2 Goodness-of-Fit of the Model

Overall, our model fits the data well. Among the 105 targeted moments, the average distance between the data moment and the corresponding model generated moment is 1.58 times the standard error of the corresponding data moment. Below we report goodness-of-fit of the model for selected moments. Appendix Tables C.VIII to C.XI report our model fits for all 105 targeted moments.

As seen in Figure 3 (a), our model can replicate the mean years smoked by age. Our model can also generate the initially high smoking initiation rate at early ages and the subsequent decline in smoking initiation rates (see Figure 3 (b)). Panel (c) shows that our model can replicate the mean years of schooling over age. Panel (d) shows that our model closely fits the mean log earnings over time. Appendix Table C.X shows that our model can replicate the median net worth at age

Table 3: Model Fit: Coefficients of Regressions for Smoking and Net Worth

Note: In Panel A, the dependent variable is the indicator variable of smoking at current period ($d_{q,t}$). The past and future smoking participation decisions $(d_{q,t-1}, d_{q,t+1})$ are instrumented using lags and leads of excise taxes and cigarette prices. In Panel B, Note: The dependent variable is the logarithm of net worth at age 30 for individuals who have at least \$100. In both panels, cognitive ability factor score and non-cognitive ability factor score are obtained from the first-stage estimation. Parameter estimate of the constant term is not a targeted moment and hence is not reported here.

	Panel A: Smoking Participation 2SLS Regression			
	Data	Model	SE of Data	$\frac{\lvert\text{Data-Model}\rvert}{\text{SE of Data}}$
Smoking at $t-1$ ($d_{q,t-1}$)	0.5415	0.7348	0.2018	0.9577
Smoking at $t+1$ ($d_{q,t+1}$)	0.3654	0.2254	0.2545	0.5502
Years of Schooling	-0.0057	-0.0027	0.0074	0.3921
Cognitive Ability	0.0010	-0.0063	0.0054	1.3545
Non-cognitive Ability	-0.0038	-0.0188	0.0095	1.5767
Log Cigarette Cost ($logp_{q,t}$)	-0.0094	-0.0050	0.0205	0.2150

	Panel B: Log Net Worth OLS Regression			
	Data	Model	SE of Data	$\frac{\lvert\text{Data-Model}\rvert}{\text{SE of Data}}$
Cognitive Ability	0.4012	0.4066	0.0625	0.0854
Non-cognitive Ability	0.2508	0.2728	0.0915	0.2402

30 and the Gini coefficient of net worth at age 30. Our model can also replicate the educational distribution for high school, some college, and 4-year college and more at age 30.

Table 3 Panel A reports smoking participation regression results for data and model simulated data (Eq. 10). The dependent variable is the indicator variable of smoking in the current period ($d_{q,t}$). Past and future smoking participation decisions $(d_{q,t-1}, d_{q,t+1})$ are instrumented using lags and leads of excise taxes and cigarette prices, following Chaloupka (1991) and Becker et al. (1994) (see the discussion in Section 2.2.2 for details regarding identification). The 2SLS regression coefficients obtained from our model simulated data fit well with the actual 2SLS regression coefficients in the data. The difference between the data and model generated moments are all well below 1.6 times standard error of the corresponding data moments.

Table 3 Panel B shows that our model replicates the coefficients of log net worth regression in the data. The model fit on all other regression coefficients including log earnings regression and schooling regression are reported in Appendix Table C.XI.

4 Policy Simulations: Effects of Education and Smoking

In this section, we conduct counterfactual simulations to evaluate the causal effects of education on smoking (Section 4.1) and smoking on education (Section 4.2). We also investigate the effects of excise taxation on smoking and education (Section 4.3). The initial distribution and number of simulations in this section are the same as in the benchmark model simulation. The total number of individuals in each simulation is 48,150 (see Section 2.3 for details).

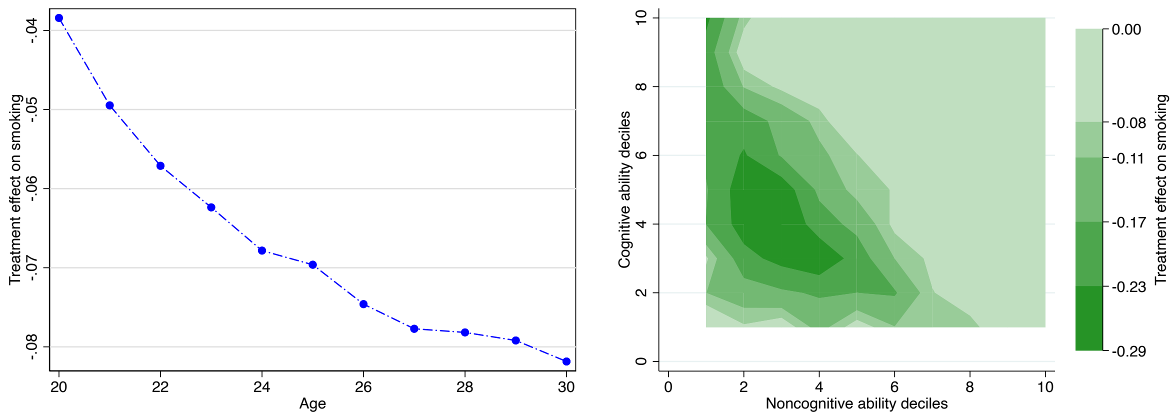
4.1 Causal Effects of Education on Smoking

To investigate the causal effects of education on smoking, we ask the following questions: (1) under autonomy (Frisch, 1938), if a youth's years of schooling is increased by one additional year, how would his smoking probability change in the current period (short-run treatment effect) and over the long run? (2) Does the treatment effect depend on the endowment distribution?

To answer these questions, we conduct counterfactual simulations where we assign each individual one extra year of schooling at age 20 compared to the benchmark model and then solve each individual's optimal choices at age 20 and onward in the counterfactual simulation.[22] The treatment effects of increasing education by one extra year are measured by the differences between the outcomes of the counterfactual simulations and the benchmark model from age 20 onward.

First, we investigate the average treatment effects of one extra year of schooling on smoking prevalence over age. The average treatment effects for a given age is measured by the changes in smoking rates between the counterfactual simulation and the benchmark model simulation, which is plotted in Figure 4 (a). One extra year of schooling at age 20 reduces the smoking rate at age 20 by 4 percentage points, a 21% reduction in comparison to the benchmark case of 18 percentage points. The effect of education is cumulative over time. Some individuals, who would have other-

[22]In this counterfactual, individuals' choices and outcomes from age 15 to age 19 are exactly the same as in the benchmark model. Starting from the beginning of age 20, we increase the years of education by one extra year in this counterfactual simulation and solve individuals' optimal decisions at age 20 and afterwards. We choose to experiment at age 20 because the smoking rate peaks at age 21 in our data and we want to investigate the effect of raising education before age 21. We have also tried to raise individuals' years of schooling at the initial age (age 15). However, over time, as individuals optimally adjust their schooling levels, their accumulated years of schooling returns to the benchmark case by age 25 in this alternative counterfactual.



(a) Change in Smoking Rate over Age (b) Change in Smoking Rate by Abilities (Age 30)

Figure 4: Causal Effects of Education on Smoking Rate

wise started smoking over time after the age of 20, avoid smoking in this counterfactual because of the additional year of education. As a result, the effects of an extra year of smoking are larger at later ages. In particular, by age-30 the smoking rate is reduced by 8 percentage points (31%) in comparison to the 27 percentage points in the benchmark case.

In this counterfactual simulation, from among the age-30 smokers in the benchmark model, 32 percent quit smoking. To understand who quits smoking in this counterfactual simulation, we conduct a regression analysis. The sample is the population of age-30 smokers in the benchmark model. Table 4 column (1) reports probit regression results. The outcome variable is one if a smoker quits smoking at age 30 in the counterfactual simulation.

Column (1) of Table 4 shows that the effects of the additional year of education accumulate over initial education. After controlling for education, the effects of an increase in cognitive ability are small and not statistically significant. Non-cognitive skills matter more than cognitive skills. One standard deviation higher non-cognitive ability increases the likelihood of quitting by 14 percentage points. This result is intuitive: when individuals with stronger skills such as self-control also receive education, they are able to better avoid behaviors such as addiction.

Finally, how does the effect of education on smoking vary by cognitive and non-cognitive ability in the entire population? Figure 4 (b) plots changes in age-30 smoking rate by cognitive ability deciles and non-cognitive ability deciles. Smokers with high cognitive and non-cognitive

Table 4: Effects on quitting smoking in the counterfactual experiments

This table reports probit analysis of quitting smoking (marginal effects), among the age-30 smokers in the benchmark model. The dependent variable is an indicator variable for not smoking at age 30 in the corresponding counterfactual simulation.

	Increasing Education	Excise Tax (40%)
	(1)	(2)
Initial years of schooling	0.145***	0.030***
	(0.004)	(0.004)
Cognitive ability	0.014	0.065***
	(0.010)	(0.004)
Non-cognitive ability	0.282***	0.122***
	(0.013)	(0.008)
Mean dependent variable	0.32	0.11
Observations	12,884	12,884
Pseudo R^2	0.238	0.053

Marginal effects; Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ability exhibit relatively less reduction in smoking as a consequence of one year of extra education. This is because these high endowment youths are less likely to smoke in the benchmark model in the first place.

However, among youths who are more likely to smoke in the benchmark model (i.e., individuals with low to medium endowments), the treatment effect of increasing education on smoking reduction becomes stronger in magnitude as we move from low ability youths to medium ability youths. The experiment thus suggests that policymakers should focus on medium ability individuals because these agents' smoking behaviors are most responsive to increases education. While cognitive and non-cognitive skills are somewhat correlated, Table 4 Column (1) further suggests that a focus on individuals with better non-cognitive skills is fruitful.

4.2 Negative Impact of Youth Smoking on Education

The literature on cigarette addiction often focuses on the negative health consequences of smoking at later ages. Not much attention has been given to investigating the potentially harmful impact of smoking and addiction at early ages from a human capital investment perspective.

Table 5: Effects on college attendance in counterfactual experiments

Panel A of this table reports the college attendance rate in different simulations. Panel B of this table reports probit regression estimates of the likelihood of college attendance in two different simulations. The dependent variable is an indicator variable for college attendance before age 30.

Panel A: College attendance rates in different simulations			
	Benchmark (1)	No Smoking (2)	Excise Tax (40%) (3)
(1) Benchmark non-college smokers	0.00	0.09	0.09
(2) Benchmark smokers	0.30	0.36	0.36
(3) All youths	0.59	0.62	0.63

Panel B: College attendance regressions			
	Benchmark (1)	No Smoking (2)	Excise Tax (40%) (3)
Cognitive Ability	0.921***	0.850***	0.748***
	(0.008)	(0.008)	(0.008)
Noncognitive Ability	0.501***	0.316***	0.380***
	(0.007)	(0.006)	(0.006)
Parents are 4-year college graduates	0.897***	0.702***	0.721***
	(0.011)	(0.009)	(0.009)
Mean dependent variable	0.59	0.62	0.63
Observations	48,150	48,150	48,150
Pseudo R^2	0.638	0.627	0.643

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To evaluate the negative impact of smoking on educational choices, we conduct a counterfactual simulation. We assign a very large negative number to the preference parameter of smoking ν_q. This approach ensures that no one in the model will find smoking attractive. The difference in college attendance rates between this counterfactual simulation (where it is never attractive to smoke) and the fitted/benchmark model simulation (where some youths obtain utility from smoking) is one measure of the loss in educational attainment due to smoking.

Table 5 reports the results. Row (1), Column (2) in Panel A shows that elimination of smoking increases the college attendance rate by 9 percentage points among the smokers in the benchmark model who would not have attended college otherwise. Driven by the increase in educational attainment among this group, college attendance rate increases by 6 percentage points among bench-

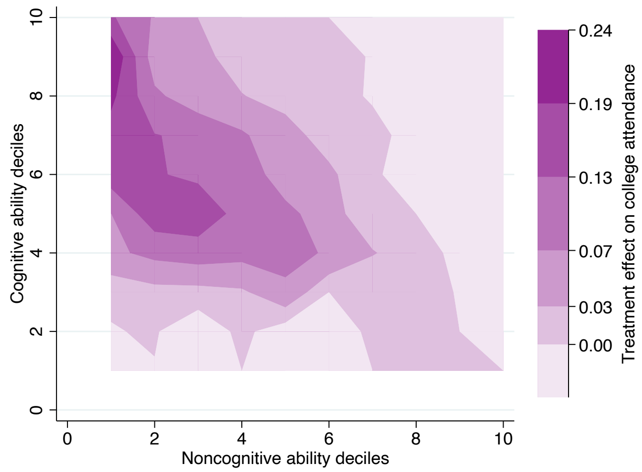


Figure 5: Causal effect of eliminating smoking on college attendance rate by ability deciles

Note: The figure plots changes in college attendance rates before age 30 between the counterfactual simulation where smoking is eliminated and the benchmark simulation.

mark model's ever smokers, from 30 percentage points to 36 percentage points in this counterfactual simulation (Table 5 Panel A Row (2), Column (2)). Over the entire population, eliminating the possibility of smoking increases college attendance rate by 3 percentage points, from 59 percentage points in the benchmark model to 62 percentage points in the counterfactual simulation (Table 5 Panel A Row (3)). Our results shed light on the negative effects of smoking on education.

Panel B of Table 5 reports probit regression estimates of likelihood of college attendance in different model simulations. We should expect that in a world without the temptation of smoking, the additional benefits of abilities such as higher self-control are reduced. This is found when we compare Columns (1) and (2) of Table 5: we find that in absence of smoking, the effect of higher non-cognitive ability on college attendance rate decreases by a third.

Comparing Tables 4 and 5 yields an additional insight: an agent's non-cognitive ability is more instrumental in avoiding harmful behavior than cognitive ability (Column 1 of Table 4). In contrast, *once harmful behavior is not possible (as in our counterfactual world)*, youths with one standard deviation higher cognitive skills attend college at four times the rate of individuals with one standard deviation higher non-cognitive skills (Column 2 of Table 5).

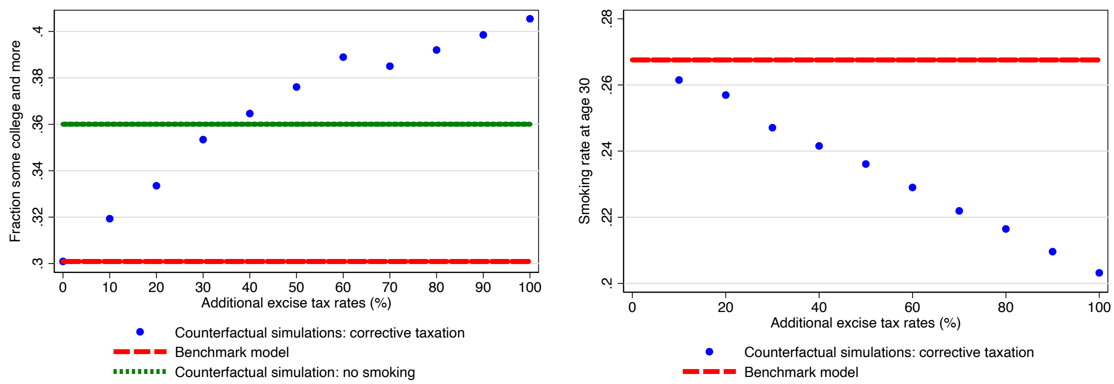


Figure 6: Effects of Corrective Excise Taxes

The left panel plots the rate of college attendance by age 30 under different counterfactual simulations among the youths who have smoked before age 30 in the benchmark model (i.e. benchmark model smokers). The right panel plots the rates of daily smoking participation at age 30 under different corrective excise tax rates. The red line indicates values in the benchmark model where there is no additional excise tax.

Finally, we document that there is substantial heterogeneity in the treatment effect of eliminating smoking on college attendance rate. Figure 5 plots the changes in college attendance (relative to the benchmark model) along the cognitive and non-cognitive ability deciles in the entire population. The increase in college attendance is larger (than the average 3 percentage points) for youths with high cognitive ability and low non-cognitive ability. These youths have low psychic cost of schooling but choose not to go to school in the benchmark model because they are addicted. In the counterfactual simulation, where smoking is eliminated, the same group of youths choose higher college attendance as the negative impact of smoking disappears.

4.3 Corrective Excise Taxes on Cigarettes

Section 4.2 shows that if smoking is made impossible, educational attainments increase. In this section, we evaluate the impact on education of a more practical policy—an excise tax on cigarettes. Such taxation does not increase smokers' perceived utility; smokers' choices are after all, rational. Nevertheless, a benevolent social planner may choose to promote flourishing lives through human capital investment. In that case, excise taxes may be attractive.

Our counterfactual experiment evaluates the effect of an additional excise tax τ on cigarettes, that is passed on to customers completely so as to increase the cost of smoking to $(1+\tau)p_{q,t}$. Note that because the cost of smoking $p_{q,t}$ in the benchmark case also includes excise taxes, the tax rate τ analyzed in this section is an additional tax on top of the current taxes. To keep the tax policy revenue neutral, we assume that the tax revenues are redistributed as a lump-sum transfer per period to the agents who attend high school (i.e., tuition subsidy). As the goal of the social planner is to improve educational outcomes, redistributing tax revenue as a tuition subsidy is more effective than redistributing tax revenue using cash transfer.

4.3.1 Effect on Education

The left panel of Figure 6 plots results of the counterfactual experiment: the college attendance rate by age 30 among the would-be smokers (in the benchmark model) at different excise tax rates. The lump-sum subsidy for a given tax rate is solved numerically through iteration so that the government's total tax revenue net the total transfer is zero for the entire population over their decision periods. As the price of cigarettes increases, many would-be smokers decide not to smoke. The reduced addiction stock increases their patience. The individuals also have more cash available as they forego cigarette purchases. These mechanisms induce these individuals to choose to go to college.[23]

The red horizontal line in the figure marks the college attendance rate (30 percent) in the benchmark model where the additional excise tax rate is zero. The green horizontal line in the figure marks the college attendance rate (36 percent) in the counterfactual simulation where smoking is eliminated (Section 4.2). As the excise tax rate increases from 10% to 100%, the college attendance rate increases from 30 percent to 41 percent. At 40% additional excise tax, the benchmark model's would-be smokers' college attendance rate reaches 36%, which is the same level of college attendance rate obtained when smoking is eliminated in the previous counterfactual experiment.

Column (3) of Panel A in Table 5 investigates who obtains additional education with 40% ex-

[23]Theoretically, excise tax alone could potentially hurt college enrollment rate. This is because smokers who chose to attend college in the benchmark case could choose not to quit smoking but rather to quit college due to the increased financial burden imposed by the excise tax.

cise tax on cigarettes. In this counterfactual, the benchmark model's non-college educated smokers increase their college attendance rate by 9 percentage points. This increase is the same as in the counterfactual simulation where smoking is eliminated as a choice. Over the entire population, 40% excise tax with tuition subsidy increases the average college attendance rate by 4 percentage points from 59 percentage points in the benchmark model to 63 percentage points. The increase in the college attendance rate is higher than the increase in college attendance when smoking is not possible. This is because some non-smokers also increase their college attendance rate due to tuition subsidy in this counterfactual simulation.

One possibility is that the excise tax impacts different smokers than the policy where smoking is not a choice. Column (3) of Table 5 Panel B checks whether the treatment effect of an excise tax is obtained from persons with different characteristics compared to the policy where smoking is eliminated as a choice. A comparison of the estimated coefficients of columns (2) and (3) assures us that the likelihoods of attending college conditional on listed characteristics are similar in both policies.

The left panel of Figure 7 plots the changes in college attendance rate after increasing the excise tax by an additional 40%, along the cognitive and non-cognitive ability deciles. Two groups of youths experience the largest improvement in their college attendance relative to the benchmark model. The first group of youths' cognitive abilities are between 4th and 8th deciles and non-cognitive abilities are below 4th deciles. The second group of youths' cognitive abilities are between 2nd and 4th deciles and non-cognitive ability is above 9 deciles.

4.3.2 Effect on Smoking

We next investigate the effects of different excise tax rate on age-30 smoking rates. The right panel of Figure 6 plots the age-30 smoking rate under different additional excise tax rates. The red horizontal line in the figure marks the age-30 smoking rate (27 percentage points) in the benchmark model. As the additional excise tax rate increases from 10% to 100%, the age-30 smoking rate decreases from 27 percentage points to 20 percentage points. At 40% additional excise tax, the

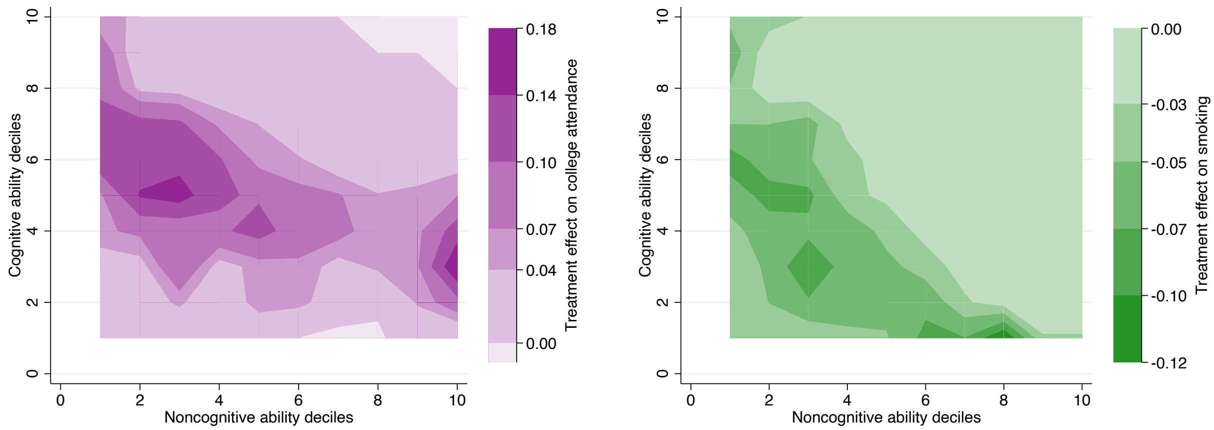


Figure 7: Causal Effects of Excise Tax by Abilities (40% additional tax)

Note: The figure plots changes in college attendance rate and changes in age-30 smoking rate under 40% additional excise tax rate.

age-30 smoking rate is reduced to 24 percentage points. The implied elasticity of daily smoking participation is -0.28, which is consistent with the estimated values in the existing studies in the smoking literature.[24]

Column (2) of Table 4 investigates which characteristics affect the quitting decision in the excise tax counterfactual among benchmark model smokers at age 30. Compared to column (1), in this counterfactual, both cognitive and non-cognitive skills determine who quits smoking. For a one standard deviation change in cognitive skills, the quitting probability increases by 4.8 percentage points. For youth with a standard deviation higher non-cognitive skills, excises taxes at 40% increase the likelihood of quitting by 6.2 percentage points.

The right panel of Figure 7 plots the changes in age-30 smoking rate with 40% additional excise tax. Youths who experience large reduction in smoking probability are in the lower triangular portion of the ability plane. Comparing with the right panel of Figure 4, the distributional effects of excise tax on smoking are similar to the distributional effects of increasing education on smoking.

[24] A review by Guindon et al. (2015) shows that the estimated price elasticity of smoking participation varies from -0.06 to -1.1. Viet Nguyen et al. (2021) finds that the price elasticity of the daily smoking prevalence is -0.26.

5 Conclusion

The decision to consume an addictive substance in the formative years of life can have long term consequences on human capital. Further education reduces addiction. To understand the reciprocal effects of addiction and education on each-other, we present and estimate a life-cycle model. The model allows youths to decide their consumption levels of an addictive good and a composite good. Youths also determine their investment in human capital while facing borrowing constraints. The choice of addiction affects flow utility and subjective discount factor. Addiction also has a monetary cost. The choice of education also affects flow utility, discount factor, available budget for consumption, and income levels.

Using our estimated model, we find that one extra year of schooling can reduce long term smoking rates by eight percentage points (31%). Non-cognitive skills are the key driver of the result: one standard deviation higher non-cognitive ability increases the likelihood of quitting by 14 percentage points. We also find that if one could eliminate smoking as a choice, college attendance rates can rise by three percentage points, from 59 percentage points to 62 percentage points in the counterfactual simulation. The higher college attendance rates can also be achieved with additional excise taxes. A 40 percentage point additional excise tax also achieves similar results. Policy impacts among smokers vary substantially by ability levels.

Our model provides a framework for investigating the origins of inequality in health behaviors and human capital and their dynamic relationships over the life-cycle. In particular, our model's framework can be applied to analyze other types of addictive behavior (e.g., marijuana and obesity) and their impacts on educational investment. Two further extensions of our model would be fruitful. First, it should be extended to study questions about inter-generational mobility and the role of harmful habits/addiction. Second, our model can be directly extended to study life-cycle inequality in health behaviors, education, and wealth by race. We leave these topics for another occasion.

References

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624.

Aboud, F. E. and Mendelson, M. J. (1996). *Determinants of friendship selection and quality: Developmental perspectives.* Cambridge University Press.

Achenbach, T. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Department of Psychiatry, University of Vermont.

Achenbach, T. and Rescorla, L. (2001). Manual for the aseba school-age forms & profiles: An integrated system of multi-informant assessment: Aseba burlington.

Achenbach, T. M. and Ruffle, T. M. (2000). The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review*, 21 8:265–71.

Adda, J. and Cornaglia, F. (2006). Taxes, cigarette consumption, and smoking intensity. *The American Economic Review*, 96(4):pp. 1013–1028.

Adda, J. and Lechene, V. (2013). Health selection and the effect of smoking on mortality. *The Scandinavian Journal of Economics*, 115(3):902–931.

Altonji, J. G. and Segal, L. M. (1996). Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14(3):353–366.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, pages 111–150, Berkeley, Calif. University of California Press.

Arcidiacono, P., Sieg, H., and Sloan, F. (2007). Living rationally under the volcano? an empirical analysis of heavy drinking and smoking. *International Economic Review*, 48(1):37–65.

Becker, G. S., Grossman, M., and Murphy, K. M. (1991). Rational addiction and the effect of price on consumption. *The American Economic Review*, 81(2):pp. 237–241.

Becker, G. S., Grossman, M., and Murphy, K. M. (1994). An empirical analysis of cigarette addiction. *The American Economic Review*, 84(3):pp. 396–418.

Becker, G. S. and Mulligan, C. B. (1997). The endogenous determination of time preference. *Quarterly Journal of Economics*, 112(3):729–758.

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, 96(4):675–700.

Becker, G. S. and Murphy, K. M. (1994). The division of labor, coordination costs, and knowledge. In Becker, G. S., editor, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, pages 299–322. University of Chicago Press, Chicago, 3 edition.

Blundell, R., Dias, M. C., Meghir, C., and Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, 84(5):1705–1753.

Browning, M., Hansen, L. P., and Heckman, J. J. (1999). Micro data and general equilibrium models. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1A, chapter 8, pages 543–633. Elsevier B. V., Amsterdam.

Burgess, S., Sanderson, E., and Umana-Aponte, M. (2011). School ties: An analysis of homophily in an adolescent friendship network. The Centre for Market and Public Organisation 11/267, The Centre for Market and Public Organisation, University of Bristol, UK.

Caucutt, E. M. and Lochner, L. (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, 128(3):1065–1147.

Centers for Disease Control and Prevention (CDC) (2010). How tobacco smoke causes disease: The biology and behavioral basis for smoking-attributable disease. A report of the surgeon general, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA.

Chaloupka, F. (1991). Rational Addictive Behavior and Cigarette Smoking. *Journal of Political Economy*, 99(4):722–42.

Chatterjee, S., Corbae, D., Nakajima, M., and Ríos-Rull, J.-V. (2007). A quantitative theory of unsecured consumer credit with risk of default. *Econometrica*, 75(6):1525–1589.

Choi, S. H. and Stommel, M. (2017). Impact of age at smoking initiation on smoking-related morbidity and all-cause mortality. *American Journal of Preventive Medicine*, 53(1):33–41.

Conti, G., Heckman, J. J., and Urzúa, S. (2010). The education-health gradient. *American Economic Review: Papers and Proceedings*, 100(2):1–5.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):pp. 87–122.

Cutler, D. M. and Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. Working Paper 12352, National Bureau of Economic Research.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behavior by education. *Journal of Health Economics*, 29:1–28.

Darden, M. (2017). Smoking, expectations, and health: A dynamic stochastic model of lifetime smoking behavior. *Journal of Political Economy*, 125(5):1465–1522.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy*, 118(1):pp. 39–75.

Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, 41(1):pp. 113–158.

Deaton, A. (2015). *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press, Princeton, New Jersey.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., and Weisz, J. R. (2011). The youth self report: Applicability and validity across younger and older youths. *Journal of Clinical Child & Adolescent Psychology*, 40(2):338–346.

Fletcher, J., Ross, S., and Zhang, Y. (2014). The Determinants and Consequences of Friendship Composition. Working Papers 2014-016, Human Capital and Economic Opportunity Working Group.

Frisch, R. (1938). Autonomy of economic relations: Statistical versus theoretical relations in economic macrodynamics. Paper given at League of Nations. Reprinted in D.F. Hendry and M.S. Morgan (1995), *The Foundations of Econometric Analysis*, Cambridge University Press.

Goldman, D. and Smith, J. P. (2011). The increasing value of education to health. *Social Science and Medicine*, 72(10):1728–1737.

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

Grossman, M. (2008). The relationship between health and schooling. *Eastern Economic Journal*, 34(3):281–292.

Gruber, J. and Köszegi, B. (2001). Is addiction "rational"? theory and evidence. *The Quarterly Journal of Economics*, 116(4):1261–1303.

Guindon, G. E., Paraje, G. R., and Chaloupka, F. J. (2015). The impact of prices and taxes on the use of tobacco products in latin america and the caribbean. *American Journal of Public Health*, 105(3):e9–e19. PMID: 25602902.

Hai, R. and Heckman, J. J. (2017). Inequality in human capital and endogenous credit constraints. *Review of Economic Dynamics*, 25:4–36. Special Issue on Human Capital and Inequality.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2016). Dynamic treatment effects. *Journal of Econometrics*, 191(2):276–292.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126(S1):S197–S246.

Heckman, J. J. and Kautz, T. (2014). Fostering and measuring skills: Interventions that improve character and cognition. In Heckman, J. J., Humphries, J. E., and Kautz, T., editors, *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, chapter 9, pages 341–430. University of Chicago Press, Chicago.

Heckman, J. J. and Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136(2):341–396.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1–2):239–267.

Heckman, J. J. and Robb, R. (1986). Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In Wainer, H., editor, *Drawing Inferences from Self-Selected Samples*, pages 63–107. Springer-Verlag, New York. Reprinted in 2000, Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, M. T. (2013). Borrowing constraints, college enrollment, and delayed entry. *Journal of Labor Economics*, 31(4):669–725.

Kaplan, G. (2012). Moving back home: Insurance against labor market risk. *Journal of Political Economy*, 120(3):446–512.

Kautz, T. and Zanoni, W. (2015). Measuring and fostering non-cognitive skills in adolescence: Evidence from Chicago Public Schools and the OneGoal Program. Unpublished manuscript, University of Chicago, Department of Economics.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

Keane, M. P. and Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103.

Kenkel, D. and Wang, P. (2001). *Rational Addiction, Occupational Choice and Human Capital Accumulation*. Academia studies in Asian economies. Edward Elgar Publishing.

Kenkel, D. S. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy*, 99(2):287–305.

Kenkel, D. S., III, R. R. R., and Wang, P. (2002). Rational Addiction, Peer Externalities and Long Run Effects of Public Policy. NBER Working Papers 9249, National Bureau of Economic Research, Inc.

Lochner, L. J. and Monge-Naranjo, A. (2011). The nature of credit constraints and human capital. *American Economic Review*, 101(6):2487–2529.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press for National Bureau of Economic Research, New York.

Navarro, S. and Zhou, J. (2017). Identifying agent's information sets: An application to a lifecycle model of schooling, consumption and labor supply. *Review of Economic Dynamics*, 25:58–92.

Orphanides, A. and Zervos, D. (1998). Myopia and addictive behaviour. *The Economic Journal*, 108(446):75–91.

Pollak, R. A. (1970). Habit formation and dynamic demand functions. *Journal of Political Economy*, 78:745 – 763.

Savelyev, P. A. and Tan, K. T. K. (2019). Socioemotional skills, education, and health-related outcomes of high-ability individuals. *American Journal of Health Economics*, 5(2):250–280.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, 28(1):78–106.

Viet Nguyen, C., Le, T. T., and Nguyen, N. H. (2021). The impact of cigarette prices on smoking participation and tobacco expenditure in vietnam. *PLOS ONE*, 16(12):1–20.

Williams, B. (2018). Identification of the Linear Factor Model. Working Papers 2018-002, The George Washington University, Department of Economics, Research Program on Forecasting.