Discussion Paper No. 600

MEASURING PEER EFFECTS ON YOUTH SMOKING BEHAVIOR

Ryo Nakajima

The Sixth ISER-Moriguchi Prize (2004)
Awarded Paper

March 2004

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Measuring Peer Effects on Youth Smoking Behavior

Ryo Nakajima *
New York University
(March 2004)

Abstract

This paper examines the role of peer effects in smoking behavior using data of middle and high school students in the United States. I present a random utility model that explicitly incorporates complementarity between individual and peer smokings. A Markov process model of smoking interactions between individuals is presented, under the assumption that such interactions occur frequently. I estimate the structural parameters of the model using a steady state distribution that is uniquely determined by the Markov process . The empirical results strongly support the presence of positive peer effects in smoking behavior among young people. Interestingly, peer interactions are found to be stronger within the same gender than across genders. The same result is found for race. Moreover, a multiplier effect is found. The impact of a tax on youth smoking increases by a factor of 1.5 when peer interactions are present.

JFL Classification: C15, C35, C63, I12

Correspondence: Ryo Nakajima, Department of Economics, New York University, 269 Mercer Street 7th Floor New York, NY 10003; e-mail: rn231@nyu.edu

*I would like to thank my thesis advisor Christopher Flinn for his guidance and supervision. I also would like to thank for valuable comments from Donghoon Lee and Giorgio Topa and helpful comments received from Victor Aguirregabiria, Robert Axtell, Alberto Bisin, Nobuyuki Hanaki, Daiji Kawaguchi, Tsunao Okumura and participants at the 2003 CIEF Conference (North Carolina), the 2003 JEA meeting (Tokyo) and seminars at Sophia University (Tokyo), Yokohama National University (Yokohama) and Institute of Social and Economic Research, Osaka University. All errors are my own.

1. INTRODUCTION

During the last three decades, federal and local government tobacco policies have achieved a dramatic reduction in the number of adult smokers in the United States. The same progress, however, has not been made for American youth. For example, the proportion of teenage smokers increased from 28 percent to 37 percent between 1992 and 1997 in the United States (Johnston et al. 2001).

A striking pattern in youth smoking data is the large differences between gender and race groups. Summarizing the four national surveys undertaken, the *Report of the Surgeon General* concludes that trends in youth smoking prevalence among gender and race groups have moved in very different directions during the last 30 years(U.S. Department of Health and Human Services 1994).

In a series of econometric studies, smoking demand functions were estimated for young people in an attempt to explain the observed differences in youth smoking behavior between groups. These studies found significant variation in price responsiveness across groups: young men and blacks are more responsive to cigarette price changes than are young women and whites.[1] However, these results raise the question of why gender and race affect the elasticity of demand for smoking.

One explanation is that the underlying cause of these difference is the intensity of peer interactions. DeCicca et al. (2000) hypothesize that peer interactions can lead to "bandwagon effects" (Liebenstein 1950) on consumption, which raise the demand for cigarettes when others are smoking. Changes in cigarette prices have not only a direct effect on consumption, but also an indirect one—changes in the consumption level of the peer group as a whole affect the demand of individual group members. Thus, differences in the intensity of such interactions could account for differences in the price elasticity between groups.[2].

In this paper, I investigate the importance of peer interactions in youth smoking behavior. The hypothesis to be tested is that the probability that an individual smokes is positively related to the fraction of smokers in his or her peer group. Data from the 2000 National Youth Tobacco Survey (NYTS) are used to test this hypothesis. This survey contains information on the prevalence of a variety of tobacco products among middle and high school students in the United States.

There are two empirical problems in estimating the magnitude of peer interactions.

The first problem is that regressing a person's behavior on the behavior of his or her peers is inappropriate. This would seem to be a natural way to estimate peer effects. However, as argued in Case and Katz (1991), peer choice is endogenous. This endogeneity creates a simultaneity problem. Peer choice, which enters a person's utility function, is also affected by that person's choice. Thus, the regression would have an error term that is correlated with peer choice, which is an explanatory variable. Standard econometric theory states that estimates from the regression would be biased and inconsistent as a result.

[1] See Chaloupka and Pacula (1999),Gruber (2000), and the *Report of the Surgeon General*(1998).

[2] Similar consumption externalities have been found by Becker (1991).

The second problem is that outcomes that are due to other factors are mistakenly attributed to peer effects. Manski (1993) argues that it is possible that peer effects may be indistinguishable from such omitted factors. An example may clarify this point. Suppose that there is a high smoking rate among teenagers in a neighborhood. This may be because they face the same cultural attitudes towards tobacco in the neighborhood, or because they have similar backgrounds as a result of their choices about where to live. One might see this as evidence of peer effects because each person's smoking seems to be due to smoking by others in the neighborhood. However, peer effects are absent because all smoking behavior in the neighborhood is due to other common factors. Failure to control for these effects may bias the estimation of peer effects.

This paper develops a method for estimating peer effects that accounts for both simultaneity and unobserved common factors. Simultaneity is addressed by treating peer outcomes as an endogenous variable that is determined by the model itself. The joint distribution of outcomes is considered in terms of probabilities. Unobserved common factors are dealt with by including fixed effects in the model. These fixed effects represent unobserved common factors that affect all members in a neighborhood.

I start with a simple dynamic model of peer interactions that determines the smoking behavior of young people. A Markov process model of smoking interactions between individuals is presented, under the assumption that such interactions occur frequently. Treating current smokers in their peer group as exogenous, individuals' conditional smoking choices describe transition probabilities.

I estimate the structural parameters of the model using a steady-state distribution that is uniquely determined by the Markov process. The distribution arises from below through repeated interactions between individuals. The steady-state distribution is assumed to specify the cross-sectional distribution of smoking profiles at any point in time. Hence, it is used to formulate the likelihood function. The structural parameters of the model are estimated using the maximum likelihood method.

Two sets of maximum likelihood estimates are obtained. The first controls for a variety of individual and county characteristics, which are augmented by the *Census* data. Since the likelihood function is intractable, it is approximated using a simulation method. Simulated samples are drawn from the Markov process introduced to describe smoking interactions. This technique was developed by Geyer and Thompson (1992). The second set of maximum likelihood estimates incorporates fixed effects into the model to account for unobserved common factors specific to neighborhoods. I suggest that these fixed effects represent unobserved neighborhood-related factors. The drawback of this approach is the associated increase in nuisance parameters. This problem is solved by using the conditional maximum likelihood method proposed by Andersen (1970).

In this paper, I focus on school cohorts as an approximate definition of peer groups. Because the data set used in this paper does not include information about the structure of peer group relations, it is necessary to make assumptions about the composition of a person's peer group. I assume that smoking interactions occur mainly between people at the same school. Since the NYTS uses samples in each school of students taking the same compulsory courses, these samples comprise students who probably see, study and play with each

other every day. Thus, this assumption is realistic. Moreover, I argue that peer interactions relate to gender and race. Thus, I estimate both gender-specific and race-specific peer effects on youth smoking behavior.

The empirical results provide compelling evidence for the existence of peer effects on young people's smoking behavior. The estimates show that peer effects are positive and highly significant. Furthermore, peer interactions are found to be stronger within genders than between genders. The same result is found for race. Furthermore, these strong peer effects are robust to the inclusion of county-specific fixed effects. These findings support the hypothesis that youth smoking patterns are due to peer effects rather than unobserved neighborhood characteristics.

The paper also examines the expected response of youth smoking behavior to changes in smoking policies. Policy experiments based on the model estimates show that a 10 percent increase in the tax on cigarettes could reduce the youth smoking rate by about 2 percent. Moreover, a multiplier effect is found. The impact of a tax on youth smoking increases by a factor of 1.5 when peer interactions are present.

The paper is organized as follows. In section 2, I present the behavior model of smoking interactions and describe the basic assumptions of the model. In section 3, I describe the empirical specification of the model and the estimation technique. In section 4, I describe the data set and provide descriptive information on the variables used for estimation. In section 5, I report the estimation results, including those from the fixed-effects model used to account for unobserved heterogeneity between counties. In section 5, I also report the results of policy experiments. Section 6 concludes the paper. Proofs and derivations of some ancillary results are presented in the appendices.

2. MODEL

2.1. *Behavioral Model*

I construct a simple interaction model based on the social capital hypothesis of (Becker 1992) to describe youth smoking behavior. The critical feature is that a person's chance of becoming addicted to smoking increases with exposure to the "social capital," which is a stock of influences from other persons. Although several alternative models are possible,[3] I adopt a simple framework by using a parameterization of the random-utility model proposed by Brock and Durlauf (2000).

Suppose that there are N persons. Persons are indexed by $i \in I \equiv (1, \cdots, N)$. Suppose person $i \in I$ is deciding whether to smoke cigarettes. The smoking state of person i is denoted by y_i, which takes the form of the following binary choice:

$$y_i = \begin{cases} +1 & \text{if smoking,} \\ -1 & \text{otherwise.} \end{cases}$$

Let $\mathbf{y}$ denote a smoking profile, which is an $N \times 1$ vector of smoking states for all N persons. That is, $\mathbf{y} = (y_1, \cdots, y_N)'$. Let Ω represent all possible states of $\mathbf{y}$; $\mathbf{y} \in \Omega$.

[3] See Glaeser and Scheinkman (2000) for various social interaction models.

For person $i \in I$, the latent utility from smoking is given by

$$y_i^* = b_i(\mathbf{x}_i) + \sum_{j \neq i} \rho_{ij} y_j + \epsilon_i. \tag{1}$$

There are two components to utility. The first utility component incorporates systematic utility ($b_i(\mathbf{x}_i)$) and a stochastic idiosyncratic taste shock (ϵ_i). In what follows, let $\mathbf{x}_i \subset \mathbb{R}^K$ be a $1 \times K$ vector of individual characteristics for person i, and let $\epsilon_i \in \mathbb{R}$ be a random taste shock for person i with a distribution $F(\epsilon)$. I assume that the variable $\mathbf{x}_i$ is observable by everyone, but that the variable ϵ_i is private information known only by person i.[4] The second utility component involves the social capital ($\sum \rho_{ij} y_j$), which is the sum of smoking outcomes of other persons. The parameter ρ_{ij} measures conformity; i.e., the degree to which person i behaves like person j. If $\rho_{ij} > 0$, person i's smoking utility y_i^* is higher when person j smokes ($y_j = +1$) than when person j does not smoke ($y_j = -1$). In short, the parameter ρ_{ij} represents the *peer effect* between person i and j.

As in the standard utility-maximization framework, decisions are made to maximize utility. While smoking utility is given by Equation (1), non-smoking utility is normalized to zero. An individual with positive latent utility chooses to smoke; that is, person i chooses smoking, $y_i = +1$, if $y_i^* \geq 0$, and chooses non-smoking, $y_i = -1$, if $y_i^* < 0$. Let y_{-i} be a smoking profile comprising the smoking states of the $(N-1)$ persons other than person i, so that $y_{-i} \equiv (y_j, j \in I \backslash \{i\})'$. Then the probability that person i smokes, ($y_i = +1$), conditional on $\mathbf{x}_i$ and y_{-i}, is given by

$$\pi_i(y_i = +1 | \mathbf{x}_i, y_{-i}) = \int_{\{\epsilon_i \in \mathbb{R} | y_i^* > 0\}} f(\epsilon_i) d\epsilon_i, \tag{2}$$

where f denotes the density function of the error distribution F.

2.2. *A Stochastic Process*

In this section, I develop a stochastic process in which each person continually updates his or her smoking status. I assume that smoking decisions are not once-and-for-all events. Many studies in developmental psychology (e.g., Flay et al. (1983)) argue that most teenagers repeat experimentation with smoking sufficiently often to acquire the smoking habit. Each person's smoking state develops in discrete steps, and it is therefore convenient to use discrete time, $t = 0, 1, 2, \cdots \in \mathbb{Z}$. Let y_i^t and $\mathbf{x}_i^t$ denote the smoking state and a vector of characteristics of person i at time t.

The specification of the stochastic process relies heavily on local interaction models of learning and adaptive behavior in game theory (e.g., Blume (1993) and Ellisson (1993)). The key features are *inertia* and *adaptive* behavior.

Inertia implies that once a decision is made, it defines behavior for some time. Suppose that each person makes a decision at randomly chosen intervals.[5] In a sufficiently small interval of time, it is unlikely that two or more persons will make decisions simultaneously.

[4] For example, a random variable ϵ_i could be interpreted as exposure to psychological stress that may lead to the onset of smoking.

[5] For example, the timing of decisions could be, but need not be, described by a Poisson process.

Let $d^t \in I$ be the person who makes a decision at time t. Then a sequence of decisions is denoted by $(d^0, d^1, d^2, \cdots)$.

Adaptive behavior implies that a person makes a decision by considering the current, not expected future, rewards of each choice. Let y_i^{*t} be the latent utility from smoking of person i at time t. Then, analogous to the latent-utility model (Equation (1)), I assume that

$$y_i^{*t} = b_i(\mathbf{x}_i^t) + \sum_{j \neq i} \rho_{ij} y_j^{t-1} + \epsilon_i^t. \tag{3}$$

In the present context, adaptive behavior means that person i at time t chooses between $y_i^t = +1$ if $y_i^{*t} \geq 0$ and $y_i^t = -1$ if $y_i^{*t} < 0$, treating other persons' choices y_j^{t-1} as exogenous. In other words, each person updates his or her choice by responding to the decisions of others observed in the previous period.

The smoking decision described by Equation (3) does not incorporate addiction. The rational addiction model (e.g., Becker and Murphy (1988)) implies that a person's current smoking consumption is determined by his or her past and expected future smoking consumption. Although addiction is a determinant of adult smoking, there are two reasons why it might not be important for young people. First, as Chaloupka (1991) shows empirically, young people tend to have higher rates of time preference for future smoking decisions than do older people. Therefore, it might be reasonable to assume that youths' current smoking demand is not influenced by expected future smoking consumption. Second, as explained in section 4, data show that young people have low levels of past smoking consumption. Therefore, one can assume that the stocks of addictive capital have negligible effects on youths' current smoking consumption. In future research, an empirical model of peer interactions with addiction and rational expectations is estimated.[6] The simpler specification, without the stock of addiction, is adopted because it seems appropriate for a study of youth smoking decisions.

To specify the probability structure of the model, the timing of decisions is crucial. If decisions occur *simultaneously*, the discrete-choice model presented above yields multiple equilibria (see Heckman (1978), Bresnahan and Reiss (1990) and Tamer (2003)). A model with multiple equilibria has neither a unique reduced form probability nor a well-defined joint probability distribution. In this case, estimation by maximum likelihood is not feasible.

The specific way in which I model smoking interactions is through a discrete-time Markov process in which each person updates his or her smoking choice sequentially over time. Markov models are often used to study complex interactions between economic agents (e.g., Föllmer (1974); Blume (1993); Ellisson (1993); Young (1993); see also Topa (2001) for an empirical application). Let $\mathbf{y}^t \equiv (y_1^t, \cdots, y_N^t)' \in \Omega$ be a smoking profile at time t, and let $\mathbf{x}^t \equiv (\mathbf{x}_1^t, \cdots, \mathbf{x}_N^t)' \in \mathbb{R}^N \times \mathbb{R}^K$ be background characteristics at time t. Consider a sequence, $\mathbf{y}^0, \mathbf{y}^1, \mathbf{y}^2, \cdots$. The transition from one state to another is given as follows. At time t, person d_t makes a decision given others' choices in the previous period. Since Equation (3) represents latent utility, the conditional probability π_i (i.e., Equation (2)) specifies the person's stochastic choice, while other persons' choices remain unchanged. Thus, a new

[6] For example, Bisin et al. (2002) study rational expectations equilibria of a model with peer interactions and incomplete information.

profile $\mathbf{y}^{t+1}$ evolves from $\mathbf{y}^t = \boldsymbol{\omega} \equiv (\omega_1, \cdots, \omega_N)$ according to the following transition: for $\xi \in \{-1, +1\}$,

$$y_i^{t+1} = \begin{cases} \xi & \text{if } i = d^t, \\ \omega_i & \text{if } i \neq d^t, \end{cases}$$

with $\text{Prob}(y_i^{t+1} = \xi | \mathbf{x}^t) = \pi_i(y_i = \xi | \mathbf{x}_i^t, y_{-i} = \omega_{-i})$. I refer to the stochastic process of a sequence of smoking profiles that is defined by the transition probability above as an *interaction process*. Because the transition probability at time $t+1$ is independent of its history before time t, the interaction process follows a Markov process on a finite state space.

I make four assumptions, which are described below. The first assumption is that peer effects are symmetric between persons. The second assumption concerns the error distribution. The third assumption requires that smoking decisions change more frequently than youth background characteristics. The fourth assumption requires that every person be able to make a decision in each time period.

Assumption 1. Peer effects are symmetric $\rho_{ij} = \rho_{ji}$ between any two persons $i, j \in I$.

Assumption 2. A shock is independent between persons and over time, is identically distributed, and has the logistic distribution $F(\epsilon_i^t) = \exp(\epsilon_i^t) \,/ [1 + \exp(\epsilon_i^t)]$.

Assumption 3. The vector of characteristics is time invariant: $\mathbf{x}^t = \mathbf{x}$ for any period $t = 0, 1, 2, \cdots$.

Assumption 4. $\text{Prob}(d^t = i) > 0$ for any person $i \in I$ and any period $t = 0, 1, 2, \cdots$.

The following results describe the properties of the *steady state* of the interaction process. (Proofs are presented in the appendix.)

Theorem 1. *Let assumptions 1–4 hold. (i) The interaction process has a unique steady-state distribution P^* such that for any $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^0 \in \Omega$*

$$\lim_{t \to \infty} \text{Prob}(\mathbf{y}^t = \boldsymbol{\omega} | \mathbf{y}^0 = \boldsymbol{\omega}^0, \mathbf{x}^0 = \mathbf{x}) = P^*(\mathbf{y} = \boldsymbol{\omega} | \mathbf{x}).$$

(ii) The steady-state distribution P^* is given by

$$P^*(\mathbf{y} = \boldsymbol{\omega} | \mathbf{x}) = \exp Q(\boldsymbol{\omega} | \mathbf{x}) \Big/ \sum_{\boldsymbol{\eta} \in \Omega} \exp Q(\boldsymbol{\eta} | \mathbf{x}), \tag{4}$$

where

$$Q(\boldsymbol{\omega} | \mathbf{x}) = \frac{1}{2} \sum_i \omega_i b_i(\mathbf{x}_i) + \frac{1}{2} \sum_{i<j} \rho_{ij} \omega_i \omega_j, \tag{5}$$

for $\boldsymbol{\omega} \in \Omega$.

The main implication of this result is that the sequence of smoking profiles $\mathbf{y}^0, \mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^T$ converges to a unique distribution P^* as T becomes large, whatever its initial distribution.

It is important to note that the model does not have multiple equilibria. The steady-state distribution P^* represents the proportion of time that the process spends in each state. Although some states of the smoking profile may occur more often than others, in principle, the distribution P^* allows each state to be revisited. Thus, the stochastic process does not converge to a few distinct equilibria.

3. EMPIRICAL STRATEGY

3.1. *Maximum Likelihood*

In this section, I discuss the estimation of the behavior model. The basic idea is that an observed smoking profile can be taken as a realization of the steady-state distribution P^*. Given data on the smoking choices y_i and background characteristics $\mathbf{x}_i$ of all $i \in N$ persons, the steady-state distribution P^* described above can be used as a likelihood function to estimate the model cross-sectionally.

In the absence of information about the structure of a person's peer group, the empirical specification assumes that a person's *school* cohort represents a well-defined peer group. Evidence from the sociology and social psychology literature indicates that the majority of middle and high school students choose as their peers fellow students from the same schools.[7] I assume that a person interacts daily with others in the same school. In other words, smoking interactions occur within, not between, schools.

Suppose that there are S different schools. Each person attends one of the finite number of non-overlapping schools, which are indexed by $s \in \{1, 2, \cdots, S\}$. Let $I_s \subset I$ denote a set of persons at school s and $N_s \equiv |I_s|$. Let $\mathbf{y}_s \equiv (y_i, i \in I_s)'$ be a smoking profile at school s, which is a vector of smoking states of N_s persons, and let $\mathbf{x}_s \equiv (\mathbf{x}_i, i \in I_s)'$ be an $N_s \times K$ matrix of individual characteristics of N_s persons at school s. In what follows, I use Ω_s to denote all possible states of $\mathbf{y}_s$. Then $\mathbf{y}_s \in \Omega_s$.

To apply the model to the observed data $(\mathbf{y}_s, \mathbf{x}_s)$ for each school $s \in \{1, 2, \cdots, S\}$, I consider the following assumption on the data generation process. Let $P_s \equiv \{\text{Prob}(\mathbf{y}_s = \boldsymbol{\omega}_s | \mathbf{x}_s) : \boldsymbol{\omega}_s \in \Omega_s\}$ be the distribution of $\mathbf{y}_s$ conditional on $\mathbf{x}_s$ in the population under consideration. For any $s \in \{1, \cdots, S\}$, $P_s = P^*$, where P^* is the steady-state distribution of the interaction process described above. Simply put, this assumes that the smoking profile observed at each school is distributed according to this steady-state distribution P^* that describes the smoking interaction process.

The objective is to estimate the structural parameters $(b_i, \rho_{ij} : i, j \in I)$ of the latent-utility model (Equation (1)). However, identification of the parameters requires the imposition of restrictions. The first identifying restriction concerns the parameterization of the systematic part of the utility function.

Assumption 5. Perceived benefits from smoking are a linear combination of background characteristics. For any $i \in I$, $b_i(\mathbf{x}_i) = \alpha + \mathbf{x}_i \boldsymbol{\beta}$, where α is a scalar and $\boldsymbol{\beta}$ is a $1 \times K$ parameter vector.

[7] Shrum et al. (1988), based on studies of friendship structure of students from grades three to 12 in a 1981–1982 survey, report that more than 95 percent of friendship links are within the same school.

The second assumption is that peer interactions depend on the *types* of person who match. Suppose that there are G different types of person, which are indexed by $g \in \{1, 2, \cdots, G\}$. Let I_g denote a set of persons of type g. Then, one can state the following.

Assumption 6. Interactions are uniform for each type. That is, between person i of type g and person j of type g' in school s, peer effects are defined by $\rho_{ij} = \rho_{gg'}/N_s$.

Peer effects are weighted by the number of persons in school s, which is N_s, so that they are independent of school size. Note that assumption 1 (i.e., $\rho_{ij} = \rho_{ji}$) implies that peer effects are symmetric between groups, which in the present context implies $\rho_{gg'} = \rho_{g'g}$ for all $g, g' \in \{1, \cdots G\}$.

As an example, I consider the *gender* of students. The most consistent finding of the literature on peer groups (e.g., Shrum et al. (1988) and McPherson et al. (2001)) is that students tend to choose school friends of the same gender. Let $g \in \{M, F\}$, where M and F represent *male* and *female* respectively. Then, gender determines within-gender and between-gender peer effects. I use ρ_{MM} to denote peer effects between a pair of persons of type M, and ρ_{MF} to denote peer effects between a pair of persons of type M and type F. The terms ρ_{FF} and ρ_{FM} are defined analogously. Since $\rho_{MF} = \rho_{FM}$ by the assumption of symmetric peer effects, the identified parameters are $(\rho_{MM}, \rho_{MF}, \rho_{FF})$, or $(\rho_{MM}, \rho_{FM}, \rho_{FF})$.

Given these assumptions, the likelihood of a smoking profile $\mathbf{y}_s$ can be defined. Let $\boldsymbol{\rho} \equiv \{\rho_{gg'}\}$ be a vector of peer effects involving $\rho_{gg'}$s for all $g, g' \in \{1, \cdots, G\}$. Then the structural parameters to be estimated are $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\rho}) \in \Theta \subset \mathbb{R}^{1+K+G^2}$. Because the steady-state distribution is given by P^* (Equation (4)), the likelihood of $\mathbf{y}_s$ conditional on $\mathbf{x}_s$ at school s is given by

$$P^*(\mathbf{y}_s = \boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) = \exp Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) \Big/ \sum_{\boldsymbol{\eta}_s \in \Omega_s} \exp Q(\boldsymbol{\eta}_s | \mathbf{x}_s, \boldsymbol{\theta}). \tag{6}$$

By analogy to Equation (5),[8] one obtains:

$$Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) = \frac{1}{2} \sum_{i \in I_s} \omega_i (\alpha + \mathbf{x}_i \boldsymbol{\beta}) + \frac{1}{2} \sum_{i < j \in I_s} \rho_{ij} \omega_i \omega_j. \tag{7}$$

The structural parameter $\boldsymbol{\theta}$ can be estimated by maximum likelihood. The contribution to the log likelihood by school s is given by

$$\ell_s(\boldsymbol{\theta}) = Q(\mathbf{y}_s | \mathbf{x}_s, \boldsymbol{\theta}) - \log \sum_{\boldsymbol{\eta}_s \in \Omega_s} \exp Q(\boldsymbol{\eta}_s | \mathbf{x}_s, \boldsymbol{\theta}). \tag{8}$$

The overall likelihood function combines the likelihood contributions of all schools; $\ell(\boldsymbol{\theta}) = \sum_s \ell_s(\boldsymbol{\theta})$. Accordingly, a maximum likelihood estimator is defined by $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$.

[8] The Q function can be also written explicitly in terms of the type-based peer effect $\rho_{gg'}$:

$$Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) = \frac{1}{2} \sum_g \sum_{I_g} \omega_i (\alpha + \mathbf{x}_i \boldsymbol{\beta}) + \frac{1}{4} \sum_g \sum_{g'} \frac{\rho_{gg'}}{N_s} \left(\sum_{I_g} \omega_i \right) \left(\sum_{I_{g'}} \omega_i \right).$$

In practice, the log likelihood function $\ell(\boldsymbol{\theta})$ requires the computation of a normalizing constant term $\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})$ (itself a function of the structural parameters). However, the exact calculation cannot be implemented analytically, and good analytical approximations are not available. Even for a moderate sample size, the computation is prohibitively expensive.[9]

I use an approach proposed by Geyer and Thompson (1992), which involves approximating the likelihood function $\ell(\boldsymbol{\theta})$ by using Monte Carlo simulations. The basic idea is to adopt an importance-sampling approach to the normalizing constant term through the use of simulated samples. Let samples $(\mathbf{y}^0,\mathbf{y}^1,\mathbf{y}^2,\cdots,\mathbf{y}^T)$ be drawn from the interaction process described in section 2, in which the structural parameter is given by $\boldsymbol{\psi}\in\Theta$. After many repetitions, T, the samples eventually converge to the steady-state distribution $P^*(\boldsymbol{\omega}|\mathbf{x},\boldsymbol{\psi})$, although they are neither independent nor distributed exactly according to $P^*(\boldsymbol{\omega}|\mathbf{x},\boldsymbol{\psi})$. These simulated samples can be used to apply the Monte Carlo method in the same way as could independent samples from the distribution $P^*(\boldsymbol{\omega}|\mathbf{x},\boldsymbol{\psi})$. Define the following function for an arbitrary fixed parameter $\boldsymbol{\psi}$:[10]

$$\ell_T(\boldsymbol{\theta};\boldsymbol{\psi}) = Q(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) - \log\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\exp Q(\mathbf{y}^t|\mathbf{x},\boldsymbol{\theta})}{\exp Q(\mathbf{y}^t|\mathbf{x},\boldsymbol{\psi})}\right], \tag{9}$$

where the school subscript s is henceforth suppressed for convenience. Given $\boldsymbol{\psi}$ and the sample size T, let $\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta}\ell_T(\boldsymbol{\theta}|\boldsymbol{\psi})$, which is known as a *Monte Carlo maximum likelihood estimator*. Geyer and Thompson (1992, pp.658–659) show, under mild conditions, that the Monte Carlo estimator $\hat{\boldsymbol{\theta}}_T$ converges almost surely to the exact estimator $\hat{\boldsymbol{\theta}}$ as $T\to\infty$, whatever $\boldsymbol{\psi}$.

---

[9] For example, if $N = 100$, it is necessary to sum over $2^{100}\simeq 10^{30}$ different states. Even a gigantic supercomputer with 100 teraflops performance, which can complete a 10^{14} summand in a second, would take a few hundred million years to compute the summation once!

[10] The maximized function $\ell_T(\boldsymbol{\theta};\boldsymbol{\psi})$ is a multiplicative constant to a Monte Carlo approximation of the log likelihood ratio of $\boldsymbol{\theta}$ against $\boldsymbol{\psi}$. That is, $\ell_T(\boldsymbol{\theta};\boldsymbol{\psi})\propto\log[\frac{P^*(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{P^*(\mathbf{y}|\mathbf{x},\boldsymbol{\psi})}]$, where

$$\begin{aligned}\log\left[\frac{P^*(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{P^*(\mathbf{y}|\mathbf{x},\boldsymbol{\psi})}\right] &= \frac{Q(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{Q(\mathbf{y}|\mathbf{x},\boldsymbol{\psi})} - \log\frac{\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})}\\ &= \frac{Q(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{Q(\mathbf{y}|\mathbf{x},\boldsymbol{\psi})} - \log\mathbf{E}\left[\frac{\exp Q(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\exp Q(\mathbf{y}|\mathbf{x},\boldsymbol{\psi})}\right].\end{aligned}$$

The last equality follows from the identity

$$\begin{aligned}\frac{\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})} &= \sum_{\boldsymbol{\eta}\in\Omega}\left[\frac{\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})}{\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})}\cdot\frac{\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})}{\sum_{\boldsymbol{\eta}\in\Omega}\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})}\right]\\ &= \sum_{\boldsymbol{\eta}\in\Omega}\left[\frac{\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})}{\exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})}P^*(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\psi})\right].\end{aligned}$$

4. DATA DESCRIPTION

The main data set used for estimation is the *2000 National Youth Tobacco Survey (NYTS)*. The survey was conducted by the American Legacy Foundation in collaboration with the CDC Foundation during the spring semester of the academic year 2000.[11] The 2000 NYTS is a school-based nationally representative sample of students from grades six to 12. Three-hundred and sixty schools were selected, and approximately five full classes in a required subject (e.g., English or Social Studies) across grades six to 12 were randomly selected from each participating school. All students in the selected classes were eligible to participate. In all, 35,828 students in 324 schools completed questionnaires. The school response rate was 90.0%, and the student response rate was 93.4%, which resulted in an overall response rate of 84.1%. Students completed an anonymous, self-administered questionnaire in the classroom, containing questions about tobacco use (bidis, cigarettes, cigars, kreteks, pipes, and smokeless tobacco), exposure to environmental tobacco smoke, the ability of minors to purchase or otherwise obtain tobacco products, knowledge of and attitudes to tobacco, and familiarity with pro- and anti-tobacco media messages.

Several sample restrictions are imposed on the data set. First, the samples are restricted to students of the four major races in the United States (whites, blacks, Hispanics and Asians). Other races (i.e., American Indians, Native Hawaiians and other Pacific Islanders) are excluded from the samples.[12] Second, the samples are also restricted to students in schools for which the 2000 *NYTS* collects at least ten male and female students.[13]. Additional restrictions due to missing observations reduce the sample size to $N = 29,385$ students and the number of schools to $S = 305$. Therefore, about 80 percent of the full sample is used for estimation. I imposed this restriction to ensure that a reasonable number of observations were available for computing meaningful average smoking outcomes for each peer group. The average sample size per school is 96.344 students with a standard deviation of 24.823 students. Maximum and minimum sample sizes are 193 and 23 respectively.

A dichotomous measure of smoking was constructed for all respondents. Each student was asked the following question: "During the past 30 days, on how many days did you smoke cigarettes?" If the answer was "none", the student was classified as a non-smoker, otherwise the student was classified as a current smoker.[14]. Table 1 reports the percentages of current smokers for gender and different races. As shown, the average proportion of current smokers among middle and high school students is 18.95 percent. Comparing smoking rates between races reveals a pattern that has also been found in recent national surveys (e.g., the *Monitoring the Future Survey*). This finding is that white youths smoke more than Hispanic youths, who in turn smoke more than black and Asian youths. Figure 1 presents the distribution of school smoking rates. Variation in smoking between schools is high: the

[11]The publicly available data set and codebook can be obtained from the website of the American Legacy Foundation (`http://www.americanlegacy.org`). See also Center for Disease Control and Prevention [CDC] (2001) for an overview of the results from this survey.

[12]This led to $1,153$ students (3.22 percent) being dropped.

[13]This led to 957 students (2.67 percent) from 13 schools being dropped.

[14]This type of smoking variable has been widely used as a smoking participation measure in previous studies of youth smoking decisions (e.g., Chaloupka and Grossman 1996, Gruber and Zinman 2000)

highest proportion of current smokers is about 60 percent and the lowest is close to zero. For compatibility with the behavior model above, the binary smoking-choice variable is set to $y_i = +1$ if person i is a current smoker, otherwise it is set to $y_i = -1$.

All respondents were also asked about the frequency and intensity of their cigarette smoking. I defined different measures of the frequency and intensity of cigarette smoking. For the frequency of smoking, six categories were defined: (1) never smoked; (2) non-smoker; (3) smoker for 1–2 days; (4) smoker for 3–9 days; (5) smoker for 10–29 days; and (6) daily smoker (all 30 days). Those who have never smoked (1) deny even trying a puff of cigarette, and non-smokers (2) have tried cigarettes but deny smoking in the past 30 days. For the intensity of smoking, six categories were defined: (1) individuals who do not smoke (no cigarettes); (2) individuals who smoke less than one cigarette per day (< 1 cigarette); (3) individuals who smoke one to five cigarettes per day (1–5 cigarettes); (4) individuals who smoke six to ten cigarettes, or half a pack, per day (6–10 cigarettes); (5) individuals who smoke 11 to 20 cigarettes, or up to one pack, per day (11–20 cigarettes); and (6) individuals who smoke more than 20 cigarettes, or more than one pack, per day (> 20 cigarettes). Table 2 presents information on the frequency and intensity of cigarette smoking.

As shown in Table 2, the students sampled did not have high levels of past cigarette consumption. For example, most students were not regular smokers: more than 95 percent of students were not daily smokers, and about 85 percent of students did not smoke at all or had smoked at least 1–2 days within the last 30 days. Smoking intensity was also low: about 90 percent of smokers smoked less than half a pack of cigarettes per day. Thus, one might reasonably characterize the students as "chippers"; i.e., just beginning to get addicted.

I have included a number of independent variables to control for factors that are thought to influence youth smoking decisions. The first set of such variables contains information on students' personal background characteristics and attitudinal attributes towards smoking. I include the following personal background characteristics that are reported in the 2000 *NYTS*: grade dummies; race dummies (for whites, blacks, Hispanics and Asians); weekly income in dollars (from paid jobs and other sources such as allowances); an employment dummy; an indicator of other smokers in the family; several indicators of school-based tobacco prevention programs (school-based program 1 practices ways to say "No" to tobacco; school-based program 2 explains why people of your age smoke; school-based Program 3 explains that most people of your age don't smoke; school-based program 4 explains the effects of smoking); and indicators of exposure to smoking on TV and in movies. To explore the effects of cigarette taxes on smoking participation, I have also included the federal and state excise taxes (in dollars) on cigarettes per pack as of 31 May 2000.[15] The data set is from Orzechowski and Walker (2001). Descriptive statistics for these control variables are presented in Table 3.

The second set of control variables reflect the characteristics of the neighborhoods in which the students live. Since the 2000 *NYTS* data do not provide any information on

[15]Cigarette taxes at 31 May are used because the 2000 *NYTS* was carried out during the spring semester. However, the estimation results are robust to changes in the period used. In fact, New York is the only state that raised the cigarette tax substantially during 2000: the state tax increased from \$0.56 to \$1.11 per pack on 1 March. Using either \$0.56 or \$1.11 as the New York state tax does not materially affect the estimation results.

the neighborhoods from which the samples are taken, I have supplemented information on county characteristics by other data sources. By exploiting information on counties from the *2000 Census* (Summary Tape File 3A) and the *2000 Uniform Crime Reporting Program Data (County-level Detailed Arrest and Offense Data)*, I have incorporated 35 county characteristics as control variables. To control for endogenous selection into neighborhoods, I have included as many attributes as possible that may affect families' decisions to locate in a given county. All the county variables used for estimation are listed in the appendix.

5. Empirical Results

5.1. *Basic Estimation Results*

In this section, I compute maximum likelihood estimates of the structural parameters $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\rho})$ of the behavioral model of smoking decisions. The Monte Carlo technique described in section 3 is used to obtain the maximum likelihood estimates; i.e., the likelihood functions are approximated using simulated samples obtained from the interaction process introduced in section 2.[16] All estimates are calculated with a final Monte Carlo sample size of $10,000$ experiments.[17] The estimated parameters from the benchmark model of youth smoking behavior are reported in Tables 4 and 5.[18] The point estimates of the county variables are presented in the appendix.

Table 4 reports the estimated coefficients of the background characteristics. As shown, almost all variables are significant and have the expected signs. The estimates provide some support for the idea that grade, race, sex, income and work experience significantly affect the probability that a student is a current smoker. All these results are consistent with the existing literature (e.g., Gruber (2000)). It is interesting to note that the coefficients of other smokers in the family are highly significant, which suggests that youths are highly likely to smoke if their parents and siblings also smoke. The evidence that family influences are important determinants of youth smoking is also consistent with previous empirical studies (e.g., Emery et al. (2001)). For ease of exposition, I treat the model specified in column

[16] I start the interaction process from the observed smoking profile in data. Before collecting Monte Carlo samples, I discarded the first $10,000$ draws on the grounds that the steady state might not have been attained by those the draws. Arguably, $10,000$ draws is sufficient for the process to reach its steady state. Then I collected simulated samples at intervals of two spacing periods to reduce autocorrelation between samples. Hence, to obtain $T = 10,000$ simulated samples, $30,000$ draws must be made from the Markov chain. A similar sampling procedure has also been suggested by Geyer and Thompson (1992).

[17] The Monte Carlo sample size of $T = 10,000$ is chosen arbitrarily. However, the sizes of the simulations are of little practical significance. I re-estimated the model using simulations of $T = 30,000$ and $T = 50,000$. The associated estimation results were virtually identical to those obtained from simulations of $T = 10,000$.

[18] Although theoretically unimportant, a good choice of ψ enhances the performance of the maximum likelihood estimator. The estimate of ψ from the standard logit regression seems a natural choice. However, for this data set, its use often produced simulated samples that yielded a likelihood function without maxima, and so the logit estimate is not a good choice. Therefore, I use a heuristic approach that searches for ψ iteratively. To be specific, I started with ψ from the standard logit estimate. Then, the one-step Newton–Raphson procedure updated ψ. Repeated updating (about 100 times) yields the appropriate value of ψ.

(3) of Table 4 as the best-fitting model because all the individual background characteristics except the constant term are statistically significant at the 5 percent level.[19]

As far as the school-based prevention programs are concerned, the estimates in Table 4 suggest that some of them reduce youth smoking rates. Specifically, school program 1 (practice ways to say "No" to tobacco) and school program 2 (which explains why youths smoke cigarettes) significantly reduce youth smoking. The significant reduction in youth smoking due to school-based program 1 is consistent with previous research, which suggests that prevention programs that teach students how to cope with peer pressure to smoke are very effective deterrents to youth smoking.[20] Curiously, however, the coefficient on school-based program 3 (which explains that most youths do not smoke cigarettes) is positive and significant, while the coefficient of school-based program 4 (which explains the effects of smoking) is negative but insignificant. These rather surprising results may indicate that school-based programs 1 and 2 are responsible for most of the preventative effects of school-based programs.

All coefficients of cigarette taxes, which appear in the bottom rows of Table 4, indicate negative and significant effects of taxation on youth smoking; i.e., higher taxes may deter youths from smoking. These results are consistent with the findings of a number of empirical econometric studies on youth smoking (e.g., Lewit et al. (1981); Chaloupka and Grossman (1996); Gruber (2000)). To illustrate the impacts of cigarette taxes on youth smoking, later I derive the tax elasticities of the youth smoking participation rate.

The estimates of peer effects (ρ_{MM},ρ_{FF}, ρ_{MF}) are reported in Table 5. Several important results emerge. First, all estimates are positive and highly significant. The peer effects are fairly large from a policy perspective. Suppose, for example, that a student moves from a school in which none of his or her peers smoke cigarettes to one in which 50 percent of the students smoke cigarettes. My results predict that in such a case, the probability of smoking would increase by 24 percentage points (from 3 percent to 27 percent) when evaluated at the sample means of the other variables. Such an effect is substantial.

Peer effects generate social multipliers. If socioeconomic conditions change, each student's smoking behavior changes not only because of the socioeconomic change but also because the smoking behavior of the peer group changes. Thus, socioeconomic change has both direct and indirect effects on youth smoking behavior. The social multiplier, which is defined as the ratio of the total effect to the direct effect, is the factor by which the externality raises the direct effect through peer interactions. Consider the following hypothetical situation. Suppose that school-based program 1 (practice ways to say "No" to tobacco) is newly introduced to schools. A prediction based on the estimated parameters shows that the youth smoking rate would fall by 3.68 percentage points from 19.71 percent to 16.03 percent. The total reduction in youth smoking can be decomposed into the direct and indirect components. For this example, the direct effect is 2.36 percentage points while the indirect effect is 1.32 percentage points. Thus, the social multiplier effect of smoking program 1 is 1.56.

[19] The estimates of other specifications are available on request.

[20] See *Report of the Surgeon General* (1998), chapter 6 (Efforts to Prevent Tobacco Use Among Young People).

It is interesting to compare estimates of social multipliers from regression-based models in other papers with those obtained in this paper. Krauth (2001) computed social multipliers for a change in the smoking rate of high school seniors (12th graders) in response to an aggregate shock. These estimated social multipliers are between 1.77 and 2.51, depending on the data set used.[21] In Gaviria and Raphael (2002), a social multiplier is estimated for a change in drug use (cocaine or marijuana) by 10th graders in response to a change in drug use by their parents.[22] Their estimated social multiplier is 1.34. Even though modeling frameworks and data are very different, these estimates are similar to those obtained in this paper. Hence, my estimates are not unreasonably large.

Table 5 shows that the magnitudes of ρ_{MM} and ρ_{FF} are not significantly different, while the *within*-gender peer effects (ρ_{MM}, ρ_{FF}) are significantly larger than the *between*-gender peer effect (ρ_{MF}). In fact, the within-gender peer effects are more than twice as large as the between-gender peer effect. This finding suggests that peer interactions are stronger within genders than between genders. In other words, when making smoking decisions, male students seem to be more influenced by their male friends than their female friends, while female students seem to be more influenced by their female friends than their male friends. Consistent with the peer network literature in sociology (e.g., Shrum et al. (1988)), this result supports the premise that smoking interactions are gender specific.

An interesting extension is to estimate race-based peer effects. Several studies in sociology (e.g., Shrum et al. (1988)) provide evidence that peer groups are formed along racial lines. If there is race homogeneity in peer groups, one would expect peer interactions to differ between racial groups. To explore this possibility, I estimate peer effects that are assumed to be constant within racial groups but different between racial groups. Assuming that *type* g is defined by race, the following peer effects are considered: $\boldsymbol{\rho} = (\rho_{WW}, \rho_{BB}, \rho_{HH}, \rho_{WB}, \rho_{BH} and \rho_{HW})$, where the subscripts W, B and H represent whites, blacks and Hispanics respectively. This specification of peer effects implies the within-race peer effects ρ_{WW}, ρ_{BB} and ρ_{HH}, and the cross-race peer effects ρ_{WB}, ρ_{BH} and ρ_{HW}. In the estimation that follows, I ignore effects for Asian students. Since Asian students comprise less than 5 percent of the total sample, there is an insufficient number of schools with at least one Asian student. Hence, all Asian students are excluded from the samples used for estimation.

One drawback with the estimation of race-based peer effects is that this requires further subsampling of the data set. As shown in Table 6, many schools are racially segregated. The table shows that almost 100 schools (approximately 30 percent) contain students from only one racial group. Most of these schools contain only white students. These racially uniform schools cannot be used to estimate between-race peer effects. In the estimation that follows, I use only samples of students in schools in which there are more than two racial groups, each of which contains at least 10 students. With these restrictions, the sample size is $N = 13,622$ individuals and the number of schools is $S = 156$. This is about 40 percent of the original sample size.

[21] The data source used by Krauth (2001) is the Monitoring the Future (MTF) survey for the period between 1976 and 1988, and the 1993 Teenage Attitudes and Practices Survey (TAPS).

[22] Gaviria and Raphael (2002) use data obtained from the 1990 Follow-up Question of the National Education Longitudinal Study (NELS).

The estimates of the peer effects $\boldsymbol{\rho}$ are given in Table 7. Since the estimated coefficients of the background characteristics are broadly similar in sign and significance to those in Table 4, estimates of $\boldsymbol{\beta}$ are not presented. Table 7 shows that all point estimates of within-race peer effects (ρ_{WW}, ρ_{BB} and ρ_{HH}) are positive and statistically significant at the 1 percent level for all races. While peer effects between white students ρ_{WW} are of a similar magnitude to those between Hispanic students ρ_{HH}, peer effects between black students ρ_{BB} are smaller, but remain substantial. Peer effects between white and Hispanic students ρ_{WH} are positive and significant, as are those between black and Hispanic students ρ_{BH}. However, peer effects between white and black students ρ_{WB} are statistically insignificant, and not always positive. This suggests that there are no peer interactions between white and black students, or negligible ones, relating to their smoking decisions. If white and black students hardly interact with each another, as the results suggest, there would be no social multiplier between the two races. This suggests that an aggregate shock that increases the smoking rate of white (or black) students would not necessary raise the smoking rate of black (or white) students. As a result, smoking rates could differ for these two racial groups.

Finally in this section, the results in Table 7 suggest that the within-race peer effects ρ_{WW}, ρ_{BB} and ρ_{HH} are substantially larger than the between-race peer effects ρ_{WB}, ρ_{HW} and ρ_{BH}. Table 5 shows similar results for gender-based peer effects. As shown, this tendency is clear among white and Hispanic students. This is illustrated in Table 7 by substantially larger values for the within-race peer effects ρ_{WW} and ρ_{HH} than for the between-race peer effects ρ_{WB} and ρ_{HW}. This implies that white and Hispanic students take account of peer behavior among students of their own race, rather than other races, when making smoking decisions. For black students, point estimates of the peer effects show that $\rho_{BH} > \rho_{BB} > \rho_{WB}$, but the null hypothesis that ρ_{BH} is significantly larger than ρ_{HH} cannot be rejected. This suggests that when making smoking decisions, black students are equally influenced by black and Hispanic students, but are hardly influenced by white students.

5.2. *Omitted Neighborhood Characteristics*

While there is evidence of strong peer effects, there are two possible sources of omitted variables bias, as suggested by Manski (1993). The first is the environmental and institutional characteristics of a neighborhood, and the second is the shared individual characteristics of a neighborhood. In the present context, (i) students in a neighborhood may be exposed to common unobserved perceptions (or cultural attitudes) towards or against tobacco and (ii) students in a neighborhood may share unobserved predispositions that lead them to smoke as a result of their families' self-selection into the neighborhood. The effect of these omitted variables, which affect everyone in a neighborhood, may be mistaken for peer effects.

To examine the possibility of omitted variables bias, I add to the model fixed effects that explicitly account for unobserved heterogeneity between neighborhoods. Fixed-effects models have been used in recent studies of peer effects (e.g., Bertrand et al. (2000); Arcidiacono and Nicholson (2002); and Weinberg et al. (2002)). I assume that everyone living in the same county may be affected by a common unobserved factor. If the unobserved heterogeneity is constant for persons in the same county, it can be accounted for by the county-specific fixed effects.

The primary assumption of the fixed-effects model is that a county-specific factor, which is unobservable to researchers, affects all persons in the same county. So, the latent-utility model is modified as follows. Let $\ell \in \{1, \cdots, L\}$ denote a county, where L is the number of counties in the sample.[23] Let δ_ℓ represent the unobserved factor specific to county ℓ, which affects all students in the county. Denote the vector of county factors by $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_L)$. I assume that the systematic utility of person i in county ℓ is given by

$$b_i(\mathbf{x}_i) = \delta_\ell + \mathbf{x}_i\boldsymbol{\beta}. \tag{10}$$

Similarly, the latent utility of person i is given by $y_i^* = b_i(\mathbf{x}_i) + \epsilon_i$; smoking $y_i = +1$ is chosen if $y_i^* \geq 0$, and non-smoking $y_i = -1$ is chosen if $y_i^* < 0$.

In the following description of the model, the likelihood function is derived for each county. I use $\mathbf{y}_\ell$ and $\mathbf{x}_\ell$ to denote the smoking profile and background characteristics in county ℓ respectively. Let I_ℓ denote the persons in county ℓ. Then $\mathbf{y}_\ell \equiv (\mathbf{y}_i, i \in I_\ell)$ is the smoking profile for county ℓ and $\mathbf{x}_\ell \equiv (\mathbf{x}_i, i \in I_\ell)$ denotes background characteristics in county ℓ. Let Ω_ℓ denote all possible states of $\mathbf{y}_\ell$ such that $\mathbf{y}_\ell \in \Omega_\ell$. Given the parameter $\boldsymbol{\theta} = (\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\rho})$, the log likelihood function of county ℓ is given by

$$\ell_\ell(\boldsymbol{\theta}) = Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) - \log \sum_{\boldsymbol{\eta}_\ell \in \Omega_\ell} \exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) \tag{11}$$

The overall log likelihood is $\ell(\boldsymbol{\theta}) = \sum_\ell \ell_\ell(\boldsymbol{\theta})$, and the maximum likelihood estimator is $\hat{\boldsymbol{\theta}} \equiv \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$.

One problem with the maximum likelihood method described above is that it is inconsistent when $N \to \infty$. This happens if the number of peer groups is fixed. This is the well-known "incidental parameter" problem (Neyman and Scott 1948). Andersen (1970) suggests that the problem can be solved if the likelihood function is conditional on the minimum sufficient statistic for incidental parameters. In the present context, the likelihood function $\ell(\boldsymbol{\theta})$ must be conditioned on the minimum sufficient statistic for the fixed-effects parameter δ_ℓ.

The conditional log likelihood function can be derived as follows. It is straightforward to show that the minimum statistic of δ_ℓ is $\tau_\ell = \sum_{i \in I_\ell} y_i$, which is the total number of smokers in county ℓ. Define all possible states of the smoking profile given the restriction that the number of smokers is fixed at τ_ℓ as follows:

$$B_\ell \equiv \left\{ \boldsymbol{\omega}_\ell \in \Omega_\ell \middle| \sum_i \omega_i = \tau_\ell \right\}.$$

The conditional probability P^* given τ_ℓ is

$$P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}; \tau_\ell) = \exp Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) \Big/ \sum_{\boldsymbol{\eta}_\ell \in B_\ell} \exp Q_1(\boldsymbol{\eta}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}_1). \tag{12}$$

[23] There are $L = 146$ counties in the sample.

(Details of the derivation are given in the appendix.) Thus, the conditional likelihood is given by

$$\ell_{1\ell}(\boldsymbol{\theta}_1) = Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) - \log \sum_{\boldsymbol{\eta}_\ell \in B_\ell} \exp Q_1(\boldsymbol{\eta}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) \tag{13}$$

where $\mathbf{x}_{1\ell}$ is a matrix of independent variables that includes individual background characteristics (e.g., grade dummies and race dummies). In other words, $\mathbf{x}_{1\ell}$ includes neither a constant nor county-specific variables (e.g., cigarette taxes and other county attributes). $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\rho})$ and $\boldsymbol{\beta}_1$ are unknown parameters corresponding to $\mathbf{x}_{1\ell}$. In this case, it can be shown that the Q_1 function does not depend on the nuisance parameter δ_ℓ as follows:

$$Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) = \frac{1}{2}\sum_{i \in I_\ell} \omega_i \mathbf{x}_{1i}\boldsymbol{\beta}_1 + \frac{1}{2}\sum_{i<j \in I_\ell} \rho_{ij}\omega_i\omega_j. \tag{14}$$

Note that the summation in the normalizing constant is taken over by the states in the smoking profile that satisfy the restriction given by τ_ℓ.

Because of its computational complexity,[24] the conditional log likelihood function $\ell_{1\ell}(\boldsymbol{\theta}_1)$ must be calculated by using the Monte Carlo technique described in section 3. By analogy to Equation (9), I define a Monte Carlo conditional likelihood function for an arbitrary fixed parameter $\boldsymbol{\psi}_1$ as

$$\ell_{1T}(\boldsymbol{\theta}_1; \boldsymbol{\psi}_1) = Q_1(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1) - \log \left[\sum_{t=1}^{T} \frac{\exp Q_1(\mathbf{y}^t|\mathbf{x}, \boldsymbol{\theta}_1)}{\exp Q_1(\mathbf{y}^t|\mathbf{x}, \boldsymbol{\psi}_1)} \right], \tag{15}$$

where the subscript ℓ is suppressed for convenience.

The simulated samples $(\mathbf{y}^0, \mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^T)$ used to construct $\ell_{1T}(\boldsymbol{\theta})$ are realizations from a stochastic process that converges to the *conditional* distribution $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1; \tau_\ell)$, which is given by Equation (12). It should be noted, however, that the interaction process used to compute the Monte Carlo *unconditional* log likelihood function (i.e., Equation (9)) cannot be used to obtain these simulated samples. This is because it generates samples that do not converge to the *conditional* distribution $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1; \tau_\ell)$, but converge to the *unconditional* distribution $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1)$.[25] In the appendix, I present an example in which the stochastic process has the conditional distribution $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1; \tau_\ell)$ as its limiting distribution.

I estimate the structural parameter $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\rho})$ using the model with county fixed effects. Using the stochastic process described in the appendix, the estimates are obtained by using the Monte Carlo conditional maximum likelihood method described above. The estimates are obtained from a final Monte Carlo sample size of $T = 10,000$, with a spacing of two full scans between simulated samples (see footnote 16).

Table 8 reports the point estimates of the coefficients $\boldsymbol{\beta}_1$ for individual background characteristics. Compare these results with those in Table 4 in which fixed effects are not controlled

[24] Since the set B has $\binom{N}{\tau}$ distinct states, computational effort rises geometrically with the sample size N.

[25] Clearly, the Markov chain of the interaction process does not converge to Equation (12) asymptotically either. Suppose that $\mathbf{y}^t = \boldsymbol{\omega} \in B$ at time t with the restriction $\sum_i \omega_i = \tau$. Let $\mathbf{y}^{t+1} = \boldsymbol{\xi}$ be generated from the Markov chain. Suppose that $d^{t+1} = j \in N$ occurs. Then ξ_j can be different from ω_j. Thus $\sum \xi_i \neq \tau$. It follows that $\mathbf{y}^{t+1} = \boldsymbol{\xi} \notin B$ occurs with a positive probability.

for. The point estimates of the coefficients $\boldsymbol{\beta}_1$ are similar. All signs are as before, and the variables that are significant in Table 4 are also significant in Table 8.

Table 9 shows the estimated peer effects $\boldsymbol{\rho}$. These peer effects are *gender based*: $\boldsymbol{\rho} = (\rho_{MM}, \rho_{MF}, \rho_{FF})$. The magnitudes of the peer effects in Table 9 are smaller than those in Table 5. For example, when county fixed effects are accounted for, the within-gender peer effects (ρ_{MM}, ρ_{FF}) decrease from 1.67 to 1.54 and from 1.62 to 1.46 respectively, while the between-gender peer effect (ρ_{MF}) decreases from 0.77 to 0.59. These are not dramatic decreases. The results in Table 9 show that these peer effects are all statistically significant, which suggests that peer groups substantially affect individual smoking behavior.

I also estimate the *race-based* peer effects $\boldsymbol{\rho} = (\rho_{WW}, \rho_{BB}, \rho_{HH}, \rho_{WB}, \rho_{BH}, \rho_{HW})$ for white, black and Hispanic students, using the model with county fixed effects. The results are reported in Table 10. These point estimates are directly comparable with those presented in Table 7. The standard errors of these estimates are larger, which indicates that the point estimates are not as precise as those shown in Table 7. This may explain why not all raced-based peer effects are statistically significant in Table 10. The estimated peer effects based on the inclusion of the fixed effects are smaller than those in Table 7. The difference is marked for the between-race peer effects. For example, the signs of the peer effects ρ_{WB} and ρ_{BH} have changed in all specifications. However, these effects are not statistically significant because the standard errors of the estimates are large. In addition, some within-race peer effects remain positive and statistically significant. The point estimate of ρ_{WW} is statistically significant at the 1 percent level, and the point estimate of ρ_{BB} is statistically significant at the 10 percent level.

In summary, many of the estimated peer effects from the model with the fixed effects (Table 9 and 10) are qualitatively similar to those from the model without the fixed effects (Table 5 and 7). The estimation results show that the inclusion of county-specific fixed effects does not eliminate the peer effects, which suggests that the estimated peer effects are not biased by the omission of county-level unobserved factors. The evidence of strong peer effects reinforces the conclusion that peer effects are not mainly driven by unobserved county characteristics.

5.3. *Policy Experiments*

To reduce youth smoking, it is essential to know whether smoking policies are effective. In this section, I evaluate the effects of smoking policies on youth smoking behavior. To do this, I repeat experiments by simulating smoking profiles from the model using the parameter estimates from the best-fitting specification from column (3) of Tables 4 and 5. Based on 1000 experiments, I compare the average values of the simulated smoking profiles under current polices and hypothetical policies.

The first policy experiment relates to cigarette taxes. Table 11 reports the predicted smoking rates under a variety of hypothetical taxes. As expected, smoking rates decrease as cigarette taxes increase. The last row of Table 11 reports the elasticities of the smoking

participation rate with respect to the tax rate.[26] These elasticities range from 0.193 to 0.210 for the tax increases considered. This implies that a 10 percent increase in federal and state taxes reduces smoking rates among middle and high school students by roughly 2 percent.

An important question about smoking prevention policies is whether higher cigarette taxes can achieve meaningful reductions in youth smoking. For example, *Healthy People 2000* sets the objective of more than doubling average federal and state taxes on cigarettes to $2.00 per pack. According to the results in Table 11, a tax increase of 120 cents per pack (a cigarette tax of $2.10 per pack) could reduce the smoking rates of middle and high school students by 4.74 percentage points, from 18.95 percent to 14.21 percent. Although this reduction is far short of the target set by *Healthy People 2000*, which is to reduce youth smoking to half its current rate, tax plays a significant role in reducing youth smoking.

There are both direct and and indirect effects of tax on smoking. Table 12 reports the decomposition of the total tax effect into the direct effect on individual outcomes and the indirect effect, which operates through peer-group interactions.[27] The last line of Table 12 reports the estimated social multiplier, which is the factor by which the externality raises the direct effect through peer interactions. As shown, the multipliers range from 1.525 to 1.649, which imply that peer effects raise the direct tax effect on youth smoking by a factor of more than 1.5. The evidence of significant multiplier effects on the tax elasticities supports the hypothesis that peer interactions strongly affect youth smoking behavior.

The second policy experiment relates to school-based smoking prevention programs. Given that the estimates strongly suggest that school-based smoking program 1 (practice ways to say "No" to tobacco) is likely to reduce youth smoking, I assess the impact of this program on youth smoking prevalence.

Table 13 shows the results of policy experiments in which school program 1 is implemented at different levels of intensity. The baseline smoking rates are computed under the assumption that all students are exposed to the same intensity level of the school-based program, which is the average level of intensity observed in the sample. Predicted smoking rates are then computed for various hypothetical average levels of intensity of the smoking program. The implied elasticities show that a 10 percent increase in the average intensity level of the school-based smoking program leads to a 0.53–0.82 percent decrease in the youth smoking rate. This finding suggests that a school-based policy is not as effective as tax policy in discouraging youth smoking. For example, the reduction achieved by a tax increase of 20 cents requires an increase in the intensity level of the school based program of about 20 percentage points.

[26] Each tax elasticity is calculated as $(\Delta y/y)/(\Delta\tau/\tau)$ where Δy is the predicted change in the smoking rate due to the assumed tax change $\Delta\tau$, y is the average smoking rate in the baseline simulation, and τ is the average tax rate for the relevant sample.

[27] I compute the direct effects as follows. Given that $\mathbf{y}$ denotes a simulated smoking profile, I compute $\Lambda_i \equiv \Lambda(-\mathbf{x}_i\hat{\boldsymbol{\beta}} - \sum_j \hat{\rho}_{ij}y_j)$ for person i, where Λ is the c.d.f of the logistic distribution, $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$ are estimated parameters, and $\mathbf{x}_i$ is a vector of characteristics including a hypothetical tax. The predicted smoking prevalence due to the direct tax change is calculated as the average of Λ_i.

6. SUMMARY AND CONCLUSIONS

If peer behavior influences individual behavior, youth smoking can differ between groups. I have used a micro data set that records the smoking behavior of middle and high school students to examine the hypothesis that peer interactions play an important role in explaining variations in youth smoking behavior.

The model presented in this paper specifies how each person's smoking behavior is related to the smoking behavior of their peers through utility. The behavior model incorporates a utility-maximization framework using the standard parameterization of the discrete-choice literature. Smoking interactions between individuals are modeled by using a Markov process, which produces a unique cross-sectional distribution of smoking profiles. This distribution is used as a likelihood function from which to estimate the model.

The model was estimated using a maximum likelihood method. The estimates show that peer effects are positive and statistically significant, and are important determinants of youth smoking. The results are robust to the inclusion of the fixed effects that control for unobserved heterogeneity between counties; i.e., peer effects remain significant. Furthermore, peer effects generate substantial externalities that lead to a more than 1.5-fold increase in the direct tax effects on youth smoking behavior. These empirical results represent consistent evidence of peer effects on youth smoking behavior. Peer effects are so important that youths succumb to smoking because of the influence of their peers.

I suggest that the framework developed in this paper could be applied to analyses of other types of behavior associated with social interactions. Recent economic studies have incorporated social interactions on the basis that interactions directly affect preferences. The empirical approach used in this paper provides a tool for testing the existence and estimating the magnitude of social interactions in settings in which there are direct interactions between individuals. Finding evidence of social interactions would enhance the practical relevance of such models.

APPENDIX A. PROOFS AND DERIVATIONS OF EQUATIONS

A.1. *Proof of Theorem*

Amemiya (1975) shows the conditions under which a set of conditional logistic probabilities can specify the joint probability in a multivariate model. Let $z_i \in \{0,1\}$ and $\mathbf{z} \equiv (z_1, \cdots, z_N) \in \{0,1\}^N \equiv \Xi$. Consider the following conditional logistic probability of z_i given z_{-i}:

$$\text{Prob}(z_i = 1|z_{-i}) = \Lambda\left(b_i(\mathbf{x}_i) + 2\sum_{j\neq i}\rho_{ij}z_j\right), \tag{16}$$

where $\Lambda(x) = (1+\exp(-x))^{-1}$ is the c.d.f. of the logistic distribution. Amemiya (1975) shows that, under assumption 1 (i.e., $\rho_{ij} = \rho_{ji}$), the conditional logistic probabilities (Equation (16)) determine at most one joint probability: For $\boldsymbol{\xi} \in \Xi$,

$$\text{Prob}(\mathbf{z} = \boldsymbol{\xi}|\mathbf{x}) = \exp Q_\xi(\boldsymbol{\xi}|\mathbf{x}) / \sum_{\boldsymbol{\eta}\in\Xi}\exp Q_\xi(\boldsymbol{\eta}|\mathbf{x}), \tag{17}$$

where

$$Q_\xi(\boldsymbol{\xi}|\mathbf{x}) = \sum_i b_i(\mathbf{x}_i)\xi_i + 2\sum_{i<j}\rho_{ij}\xi_i\xi_j. \tag{18}$$

Making use of the above result, I show that a set of conditional probabilities π_i (Equation (2)) for $i \in I$ determines the joint probability P^* (Equation (4)). Under assumption 2 (i.e., $F(\epsilon) = \Lambda(\epsilon)$), π_i is written as

$$\pi_i(y_i = +1|y_{-i}) = \Lambda\left(b_i(\mathbf{x}_i) + \sum_{j\neq i}\rho_{ij}y_j\right).$$

Using the variable transformation $z = (y+1)/2$ reveals that the same probability mass as π_i is assigned by

$$\text{Prob}(z_i = 1|z_{-i}) = \Lambda\left(b_i(\mathbf{x}_i) + 2\sum_{j\neq i}\rho_{ij}z_j - \sum_{j\neq i}\rho_{ij}\right).$$

Therefore, by analogy to Equation (17) and Equation (18), the conditional probabilities π_i for $i \in I$ determine the joint probability

$$\text{Prob}(\mathbf{z} = \boldsymbol{\xi}|\mathbf{x}) = \exp Q_\xi(\boldsymbol{\xi}|\mathbf{x}) / \sum_{\boldsymbol{\eta}\in\Xi}\exp Q_\xi(\boldsymbol{\eta}|\mathbf{x}), \tag{19}$$

where

$$Q_\xi(\boldsymbol{\xi}|\mathbf{x}) = \sum_i\left(b_i(\mathbf{x}_i) - \sum_{j\neq i}\rho_{ij}\right)\xi_i + 2\sum_{i<j}\rho_{ij}\xi_i\xi_j. \tag{20}$$

Equation (19) can be expressed in terms of $\mathbf{y}$ using the variable transformation as

$$\text{Prob}(\mathbf{y} = \boldsymbol{\omega}|\mathbf{x}) = \text{Prob}\left(\mathbf{z} = \frac{\boldsymbol{\omega}+1}{2}\middle|\mathbf{x}\right).$$

Hence, this is equal to

$$\text{(21)} \qquad \text{Prob}(\mathbf{y}=\boldsymbol{\omega}|\mathbf{x}) = \exp Q_\xi\left(\frac{\boldsymbol{\omega}+1}{2}\middle|\mathbf{x}\right) \Big/ \sum_{\boldsymbol{\eta}\in\Xi} \exp Q_\xi\left(\frac{\boldsymbol{\eta}+1}{2}\middle|\mathbf{x}\right).$$

Equation (20) implies that

$$\begin{aligned} Q_\xi\left(\frac{\boldsymbol{\omega}+1}{2}\middle|\mathbf{x}\right) &= \sum_i \left(b_i(\mathbf{x}_i) - \sum_{j\neq i}\rho_{ij}\right)\left(\frac{\omega_i+1}{2}\right) + 2\sum_{i<j}\rho_{ij}\left(\frac{\omega_i+1}{2}\right)\left(\frac{\omega_j+1}{2}\right) \\ &= \frac{1}{2}\sum_i b_i(\mathbf{x}_i)\omega_i + \frac{1}{2}\sum_{i<j}\rho_{ij}\omega_i\omega_j + constant. \end{aligned}$$

Recall that Equation (5) is

$$Q(\boldsymbol{\omega}|\mathbf{x}) \equiv \frac{1}{2}\sum_i b_i(\mathbf{x}_i)\omega_i + \frac{1}{2}\sum_{i<j}\rho_{ij}\omega_i\omega_j.$$

Then,

$$\text{(22)} \qquad Q_\xi\left(\frac{\boldsymbol{\omega}+1}{2}\middle|\mathbf{x}\right) = Q(\boldsymbol{\omega}|\mathbf{x}) + constant.$$

Combining Equation (22) with Equation (21) yields Equation (4). Therefore, I have shown that Equation (2) determines Equation (4) such that

$$\text{(23)} \qquad P^*(y_i=\omega_i|\mathbf{x}, y_{-i}=\omega_{-i}) = \pi_i(\omega_i|\mathbf{x}_i, \omega_{-i}),$$

for any $\boldsymbol{\omega}\in\Omega$ and $i\in I$.

I now investigate the interaction process $\mathbf{y}^0$, $\mathbf{y}^1$, $\mathbf{y}^2$, $\cdots$,$\mathbf{y}^T$. The transition probability of $i=d^t$ is given by

$$\text{Prob}(y^t_{d^t}=\omega_{d^t}|\mathbf{x}^t, y^{t-1}_{-d^t}=\omega_{-d^t}) = \pi_{d^t}(\omega_{d^t}|\mathbf{x}_{d^t}, \omega_{-d^t}),$$

under assumption 3 (i.e., $\mathbf{x}^t=\mathbf{x}$ for any t). Given Equation (23), this is equal to

$$\text{Prob}(y^t_{d^t}=\omega_{d^t}|\mathbf{x}^t, y^{t-1}_{-d^t}=\omega_{-d^t}) = P^*(y^t_{d^t}=\omega_{d^t}|\mathbf{x}, y^{t-1}_{-d^t}=\omega_{-d^t}).$$

In mathematical terms,

$$\text{(24)} \qquad \text{Prob}(y^t_i=\omega_i) = P^*(y^t_{d^t}=\omega_{d^t}|\mathbf{x}, y^{t-1}_{-d^t}=\omega_{-d^t})\cdot\text{Prob}(y_{-d^t}=\omega_{-d^t}).$$

Geman and Geman (1984, Theorem A) prove that Equation (24) implies that the above Markov chain converges to P^* as $t\to\infty$ whatever the initial profile $\mathbf{y}^0$; namely,

$$\lim_{t\to\infty}\text{Prob}(\mathbf{y}^t=\boldsymbol{\omega}|\mathbf{y}^0=\boldsymbol{\omega}^0, \mathbf{x}^0=\mathbf{x}) = P^*(\mathbf{y}=\boldsymbol{\omega}|\mathbf{x}).$$

The only necessary condition for convergence is that a sequence of $(d^0, d^1, d^2, \cdots)$ contains any $i\in N$ infinitely often, which is ensured by assumption 4 (i.e., $\text{Prob}(d^t=i)>0$ for any i and t). This proves result 1. □

A.2. *Derivation of the Conditional Probability P^* on τ*

In the following derivation, I use a decomposition of the vector of characteristics $\mathbf{x}_i$. Let $\mathbf{x}_{1i}$ denote a vector of individual characteristics (e.g., grade dummies, race dummies), and let $\mathbf{x}_{2i}$ denote a vector of constant and county characteristics (e.g., cigarette tax and other county characteristics). The $\mathbf{x}_{2i}$ does not change across individuals in county ℓ. The parameters are decomposed into $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ denote vectors of unknown parameters that correspond to $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ respectively.

Using the above notation, the Q function is written as

$$Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) = \frac{1}{2}\sum_i \omega_i(\delta_\ell + \mathbf{x}_{1i}\boldsymbol{\beta}_1 + \mathbf{x}_{2i}\boldsymbol{\beta}_2) + \frac{1}{2}\sum_{i<j}\rho_{ij}\omega_i\omega_j.$$

Under the restriction that $\tau_\ell = \sum_i \omega_i$, this is equal to

$$\begin{aligned} Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) &= \frac{1}{2}\sum_i(\delta_\ell + \mathbf{x}_{2i}\boldsymbol{\beta}_2)\tau_\ell + \frac{1}{2}\sum_i \omega_i\mathbf{x}_{1i}\boldsymbol{\beta}_1 + \frac{1}{2}\sum_{i<j}\rho_{ij}\omega_i\omega_j \\ &= C_\ell + Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}), \end{aligned}$$

where I define $C_\ell \equiv \frac{1}{2}\sum_i(\delta_\ell + \mathbf{x}_{2i}\boldsymbol{\beta}_2)\tau_\ell$.

The rule of conditional probability implies

$$P^*(\mathbf{y}_\ell = \boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}; \tau_\ell) = P^*(\mathbf{y}_\ell = \boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) \Big/ \sum_{\boldsymbol{\omega}_\ell \in B_\ell} P^*(\mathbf{y}_\ell = \boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}).$$

This is equal to

$$\begin{aligned} P^*(\mathbf{y}_\ell = \boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}; \tau_\ell) &= \frac{\exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}_\ell \in \Omega_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta})} \Big/ \frac{\sum_{\boldsymbol{\omega}_\ell \in B_\ell}\exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}_\ell \in \Omega_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta})} \\ &= \exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) \Big/ \sum_{\boldsymbol{\eta}_\ell \in B_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}) \\ &= \exp C_\ell \exp Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}) \Big/ \sum_{\boldsymbol{\eta}_\ell \in B_\ell}\exp C_\ell \exp Q_1(\boldsymbol{\eta}_\ell|\mathbf{x}_{1\ell}, \boldsymbol{\theta}). \end{aligned}$$

The term $\exp C_\ell$, which is constant, is canceled out. This is equal to Equation (12). □

APPENDIX B. INCLUDED COUNTY-SPECIFIC VARIABLES

B.1. *2000 Census data*

The following county-specific variables are provided by the *2000 Census Data*, Summary Tape File 3.

Persons 0-18: The percentage of persons between the age of 0 and 18 in the county.

Persons 65+: The percentage of persons of the age of 65 and older in the county.

Asian persons: The percentage of Asian persons in the county.

Black persons: The percentage of black persons in the county.

Hispanic persons: The percentage of Hispanic persons in the county.

Single-mother families: The percentage of female householder families; no husband present; with own children under the age of 18.

One-person families: The percentage of one-person non-family households.

Persons with grandchildren: The percentage of persons of the age of 30 and older who live with their own grandchildren under the age of 18.

Persons who divorced: The percentage of persons of the age of 15 and older who had been divorced.

Unemployed workers: The percentage of unemployed persons over the civilian labor force of the age of 16 and older.

Workers with management jobs: The percentage of persons of the age of 16 and older employed in management; business; and financial operations occupations.

Workers with professional jobs: The percentage of persons of the age of 16 and older employed in professional and related occupations.

Workers with construction jobs: The percentage of persons of the age of 16 and older employed in construction; extraction; and maintenance occupations.

Workers with farming jobs: The percentage of persons of the age of 16 and older employed in farming; fishing; and forestry occupations.

Workers with sales and office jobs: The percentage of persons of the age of 16 and older employed in sales and office occupations.

Workers with service jobs: The percentage of persons of the age of 16 and older employed in service occupations.

Persons below poverty level: The percentage of persons with income below the poverty level.

Persons who graduated high schools: The percentage of persons of the age 16 and older with a high school diploma or more.

Persons who graduated colleges: The percentage of persons of the age 16 and older with a college degree or more.

Persons enrolled in public schools: The percentage of persons of the age of 3 or older who are enrolled in a public school at present.

Persons enrolled in high schools: The percentage of persons of the age of 3 or older who are enrolled in a high school at present.

Persons enrolled in colleges: The percentage of persons of the age of 3 or older who are enrolled in a college at present.

Persons enrolled in graduate schools: The percentage of persons of the age of 3 or older who are enrolled in a graduate or professional school at present.

Persons in armed forces: The percentage of persons between the age of 18 and 64 who are in armed forces.

Persons in veteran status: The percentage of persons between the age of 18 and 64 who are civilian veterans.

Persons migrated from foreign countries: The percentage of persons who were born in the countries other than United States.

Persons moved in last 5 years: The percentage of persons who moved in the county in last 5 years.

Median household income: in 1999 U.S. $, includes wage or salary income, self-employment income, interest, dividend or net rental income, social security, public assistance, retirement income, and other.

Median housing value: Median housing value in 1999 U.S. $ of owner occupied non-condominium housing units in the county.

Vacant housing units: The percentage of all housing units that are vacant in the county.

Private vehicles occupancy: The percentage of workers of the age 16 and older who owns private cars, vans and trucks.

Urban area rate: The percentage of urbanized area over total area in the county.

B.2. *2000 Uniform Crime Reporting Program Data*

The following county-specific variables are provided by the *2000 Uniform Crime Reporting Program Data: County-level Detailed Arrest and Offense Data*.

Crime index: The percentage of arrests and offenses for the Uniform Crime Reports index crimes (murder, rape, robbery, aggravated assault, burglary, larceny, auto theft and arson) over the population in the county.

Juveniles crime: The percentage of arrests and offenses of juveniles for the Uniform Crime Reports index crimes (murder, rape, robbery, aggravated assault, burglary, larceny, auto theft and arson) over the population in the county.

Juveniles crime (drug abuse): The percentage of arrests of juveniles for violations of narcotic drug over the population in the county.

APPENDIX C. AN EXAMPLE OF A MARKOV CHAIN

The following Markov chain has the conditional distribution $P^*(\boldsymbol{\omega}|\mathbf{x}, \tau, \boldsymbol{\theta})$, which is given by Equation (12), as the steady-state distribution. Given $\mathbf{x}^t = \mathbf{x}$ and $\mathbf{y}^t = \boldsymbol{\omega} \in B$ at time t, construct $\boldsymbol{\xi} \in B$ by swapping two elements of $\boldsymbol{\omega}$. For example, let i and j be the randomly selected pair of persons ($i, j \in N$). Then $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_i, \cdots, \omega_j, \cdots, \omega_N)$ generates $\boldsymbol{\xi} = (\omega_1, \cdots, \omega_j, \cdots, \omega_i, \cdots, \omega_N)$.

Consider a Markov chain in which a new profile $\mathbf{y}^{t+1}$ evolves from $\mathbf{y}^t$ according to

$$\mathbf{y}^{t+1} = \begin{cases} \boldsymbol{\omega} & \text{with probability } \alpha(\boldsymbol{\omega}, \boldsymbol{\xi}), \\ \boldsymbol{\xi} & \text{with probability } 1 - \alpha(\boldsymbol{\omega}, \boldsymbol{\xi}). \end{cases}$$

Here, I define

$$\alpha(\boldsymbol{\omega}, \boldsymbol{\xi}) = \min \left\{ \frac{Q_1(\boldsymbol{\omega}|\mathbf{x}, \boldsymbol{\theta})}{Q_1(\boldsymbol{\xi}|\mathbf{x}, \boldsymbol{\theta})}, 1 \right\}.$$

The Markov process states that the next candidate value is proposed by swapping choices between the randomly selected individuals, and is then accepted or rejected according to the potential function at the candidate value relative to the potential function at the current value. Such an algorithm is often referred to as the *Metropolis* sampling algorithm in the statistical literature. See Chib and Greenberg (1996) for an introduction to the Metropolis algorithm and its application to econometric problems.

Some important points should be noted about the above Markov process.

First, the Markov process satisfies the summation restriction imposed by B. Suppose that $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_N) \in B$ with the restriction $\sum_i \omega_i = \tau$. Since just two elements of $\boldsymbol{\omega}$ are swapped to construct $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_N)$ at the next step, it follows that $\sum_i \xi_i = \sum_i \omega_i = \tau$, and thus $\boldsymbol{\xi} \in B$. Because either $\mathbf{y}^{t+1} = \boldsymbol{\omega}$ or $\mathbf{y}^{t+1} = \boldsymbol{\xi}$, it follows that $\mathbf{y}^t \in B$ implies $\mathbf{y}^{t+1} \in B$.

Second, the Markov process is aperiodic. This is because the algorithm states that the probability that the next draw is the same as the current draw is positive. This happens when the proposed sample is rejected.

Finally, the Markov process is irreducible. To move from state $\boldsymbol{\omega}$ to state $\boldsymbol{\omega}'$, find the coordinates whose values in $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are different, and swap these two values. Continue swapping until $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ agree. If pairs swapped are chosen randomly, such a transition from state $\boldsymbol{\omega}$ to state $\boldsymbol{\omega}'$ occurs with positive probability for any $\boldsymbol{\omega}, \boldsymbol{\omega}' \in B$. That means that there is a positive probability of reaching any state from any other state in finite steps, so that the process visits all the states of B. The standard result shows that if the Markov process is aperiodic and irreducible, it converges asymptotically. In particular, a sequence $\mathbf{y}^0, \mathbf{y}^1, \mathbf{y}^2, \cdots$ converges in terms of its distribution to the conditional distribution given by Equation (12).

REFERENCES

Amemiya, Takeshi, "Qualitative Response Models," *Annals of Economics and Social Measurement*, 1975, *4*, 363–372.

Andersen, Erling B., "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of The Royal Statistical Society, Series B*, 1970, pp. 283–301.

Arcidiacono, Peter and Sean Nicholson, "Peer Effects in Medical School," 2002. NBER Working Paper #9025.

Becker, Gary S., "A Note on Restaurant Pricing and Other Example of Social Influences on Price," *Journal of Political Economy*, 1991, *99*, 1109–1116.

_____, "Habits, Addiction, and Tradition," *Kyklos*, 1992, *45*, 327–346.

_____ **and Kevin M. Murphy**, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96*, 675–700.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan, "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, 2000, *115*, 1019–1055.

Bisin, Alberto, Ulrich Horst, and Onur Özgür, "Rational Expectations Equilibria of Economies with Local Interactions," 2002. mimeo, New York University.

Blume, Lawrence E., "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 1993, *5*, 387–424.

Bresnahan, Timothy F. and Peter C. Reiss, "Entry in Monopoly Markets," *Review of Economic Studies*, 1990, *57* (4), 531–553.

Brock, William A. and Steven N. Durlauf, "Interactions-Based Models," in James J. Heckman and Edward Learner, eds., *Handbook of Econometrics*, Vol. 5 2000, chapter 54, pp. 3297–3380.

Case, Anne C. and Lawrence F. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youth," 1991. NBER Working Paper #3705.

Center for Disease Control and Prevention [CDC], "Youth Tobacco Surveillance United States – 2000," *CDC Surveillance Summaries, MMWR*, 2001, *50* (No. SS-4).

Chaloupka, Frank J., "Rational Addictive Behavior and Cigarette Smoking," *Journal of Political Economy*, 1991, *99*, 722–742.

_____ **and Michael Grossman**, "Price, Tobacco Control Policies and Youth Smoking," 1996. NBER Working Paper #5740.

_____ **and Rosalie L. Pacula**, "Sex and Race Difference in Young People's Responsiveness Price and Tobacco Control Policies," *Tobacco Control*, 1999, *8*, 373–377.

Chib, Siddhartha and Edward Greenberg, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 1996, *12*, 409–431.

DeCicca, Phillip, Donald Kenkel, and Alan Mathios, "Racial Difference in the Determinants of Smoking Onset," *Journal of Risk and Uncertainty*, 2000, *21*, 311–340.

Ellisson, Glenn, "Learning, Local Interaction, and Coordination," *Econometrica*, 1993, *61*, 1047–1071.

Emery, Sherry, Martha M. White, and John Pierce, "Does Cigarette Price Influence Adolescent Experimentation," *Journal of Health Economics*, 2001, *20*, 261–270.

Flay, Brian R., Josie R. d'Avernas, May W. Kersell, and Katherine B Ryan, "Cigarette Smoking: Why Young People Do It and Ways of Preventing It," in Patrick J. McGrath and Phillip Firestone, eds., *Pediatric and Adolescent Behavioral Medicine*, New York: Springer, 1983, chapter 6, pp. 132–181.

Föllmer, Hans, "Random Economics with Many Interaction Agents," *Journal of Mathematical Economics*, 1974, *1*, 51–62.

Gaviria, Alejandro and Steven Raphael, "School-based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, 2002, *83*, 257–268.

Geman, Stuart and Donald Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, *PAMI-6*, 721–741.

Geyer, Charles J. and Elizabeth A. Thompson, "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of The Royal Statistical Society, Series B*, 1992, *54*, 657–699.

Glaeser, Edward and Josè A. Scheinkman, "Non-Market Interactions," 2000. NBER Working Paper #8035.

Gruber, Jonathan, "Youth Smoking in the U.S.: Prices and Policies," 2000. NBER Working Paper #7506.

_____ **and Jonathan Zinman**, "Youth Smoking in the U.S.: Evidence and Implications," 2000. NBER Working Paper #7780.

Heckman, James J., "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–959.

Johnston, Lloyd D., Patrick M. O'Malley, and Jerald G. Bachman, "Monitoring The Future National Survey Results On Drug Use, 1975-2000. Volume I: Secondary Students," 2001. National Institute on Drug Abuse: NHIS Pub. No. 01-4924.

Krauth, Brian, "Peers and Prices: Explaining the Black-white Youth Smoking Gap," 2001. mimeo, Simon Fraser University.

Lewit, Eugene M, Douglas Coate, and Michael Grossman, "The Effects of Government Regulation on Teenage Smoking," *Journal of Law and Economics*, 1981, *24*, 545–569.

Liebenstein, Harvey, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumer Demand," *Quarterly Journal of Economics*, 1950, *64*, 183–207.

Manski, Charles F., "Identification of Endogenous Social Effects: the Reflection Problem," *Review of Economic Studies*, 1993, *60*, 531–542.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook, "Birds of a Feather: Homophyly in Social Networks," *Annual Review of Sociology*, 2001, *27*, 415–444.

Neyman, J. and Elizabeth L. Scott, "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 1948, *16*, 1–32.

Orzechowski, W. and R. C. Walker, *The Tax Burden on Tobacco: Historical Compilation 2000*, Arlington, VA, 2001.

Shrum, Wesley, Neil H. Cheek, and Saundra MacD. Hunter, "Friendship in School: Gender and Racial Homophyly," *Sociology of Education*, 1988, *61*, 227–239.

Tamer, Elie T., "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 2003, *70*, 147–165.

Topa, Giorgio, "Social Interaction, Local Spillovers and Unemployment," *Review of Economic Studies*, 2001, *68*, 261–295.

U.S. Department of Health and Human Services, "A Report of the Surgeon General: Preventing Tobacco Use Among Young People," 1994.

_____, "A Report of the Surgeon General: Tobacco Use among U.S. racial/ethnic minority groups – African Americans, American Indians and Alaska Natives Pacific Islanders, and Hispanic," 1998.

Weinberg, Bruce A., Patricia B. Reagan, and Jeffrey J. Yankow, "Do Neighborhoods Affect Hours Worked: Evidence from Longitudinal Data," 2002. mimeo, Department of Economics and Center for Human Resource Research Ohio State University.

Young, H. Peyton, "The Evolution of Conventions," *Econometrica*, 1993, *61* (1), 57–84.

TABLE 1. Smoking Rates (Percentages) by Gender and Race Subgroups in Sample

	Gender		Total
Race	Male	Female	
White	21.829	22.192	22.012
Black	15.499	11.153	13.158
Hispanic	18.111	13.507	15.745
Asian	17.237	11.425	14.486
Total	19.874	18.059	18.951

TABLE 2. Smoking Frequency and Intensity

Intensity/day	Frequency/30days						Sum
	Never Smokers	Non-Smokers	Smokers 1-2 days	Smokers 3-9 days	Smokers 10-29 days	Daily Smokers	
0 cig	16680	5391	0	0	0	0	22071
$<$1cig	0	0	741	209	48	4	1002
1-5cigs	0	0	608	929	919	485	2941
6-10cigs	0	0	22	63	163	518	766
11-20cigs	0	0	2	10	59	387	458
$>$ 20cigs	0	0	0	3	9	199	211
Sum	16680	5391	1373	1214	1198	1593	27449

Never Smokers: those who deny ever trying a puff or two of cigarettes.
Non-Smokers: those who endors trying cigarettes, although deny smoking within 30 days.
Daily Smokers: those who smoke everyday within 30 days.

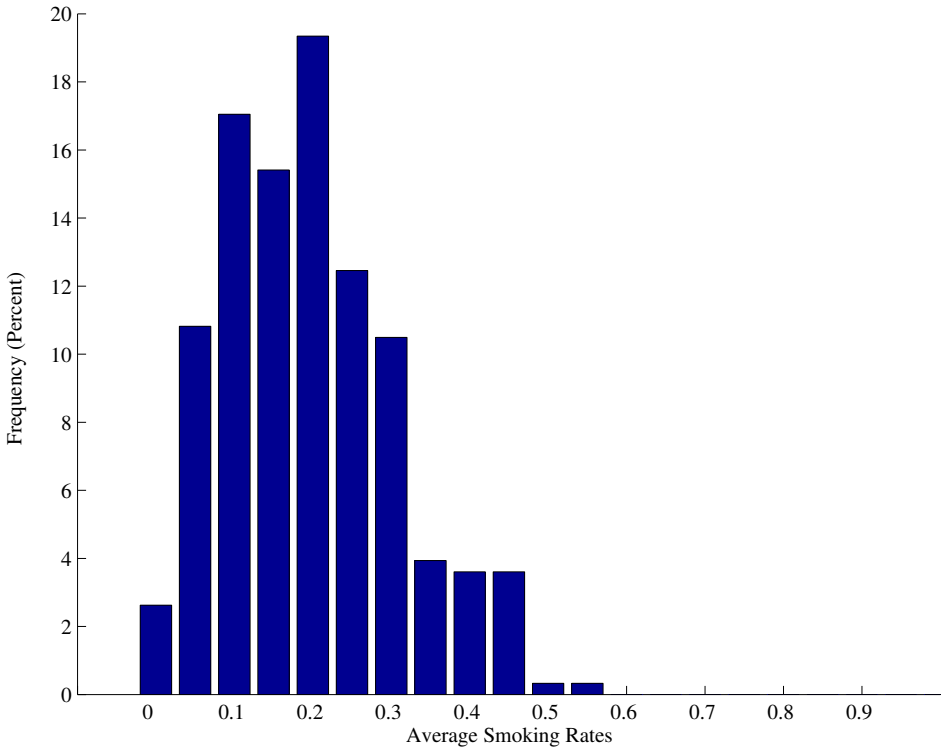

FIGURE 1. Distribution of School Smoking Rates

TABLE 3. Descriptive Statistics for Individual Background Variables Used in Estimation

Variable	Minimum	Maximum	Mean	Standard Deviation
7th Grade (0-1 Dummy)	0.00000	1.00000	0.15604	0.36290
8th Grade (0-1 Dummy)	0.00000	1.00000	0.16259	0.36900
9th Grade (0-1 Dummy)	0.00000	1.00000	0.15635	0.36319
10th Grade (0-1 Dummy)	0.00000	1.00000	0.13788	0.34478
11th Grade (0-1 Dummy)	0.00000	1.00000	0.13977	0.34676
12th Grade (0-1 Dummy)	0.00000	1.00000	0.10756	0.30983
Asian (0-1 Dummy)	0.00000	1.00000	0.04994	0.21783
Black (0-1 Dummy)	0.00000	1.00000	0.16904	0.37479
Hispanic (0-1 Dummy)	0.00000	1.00000	0.18691	0.38985
White Male (0-1 Dummy)	0.00000	1.00000	0.29381	0.45551
Asian Male (0-1 Dummy)	0.00000	1.00000	0.02646	0.16051
Black Male (0-1 Dummy)	0.00000	1.00000	0.07781	0.26787
Hispanic Male (0-1 Dummy)	0.00000	1.00000	0.09074	0.28724
Weekly Income (U.S. Dollar)	0.00000	46.71429	6.43774	8.36576
Working Dummy (0-1 Dummy)	0.00000	1.00000	0.39563	0.48899
Smokers in Home (0-1 Dummy)	0.00000	1.00000	0.40516	0.49093
See Actors Smoking in TV (0-1 Dummy)	0.00000	1.00000	0.83573	0.37053
See Actors Smoking in Movie (0-1 Dummy)	0.00000	1.00000	0.86279	0.34408
School Program 1 (0-1 Dummy)	0.00000	1.00000	0.27043	0.44419
School Program 2 (0-1 Dummy)	0.00000	1.00000	0.42268	0.49399
School Program 3 (0-1 Dummy)	0.00000	1.00000	0.22080	0.41480
School Program 4 (0-1 Dummy)	0.00000	1.00000	0.56102	0.49627
Cigarette Tax (U.S. Dollar)	0.36500	1.45000	0.89206	0.32863

School Program 1: Practice ways to say “No” to tobacco

School Program 2: Taught reasons why people your age smoke

School Program 3: Taught that most people your age don’t smoke

School Program 4: Taught about the effects of smoking

TABLE 4. The estimated coefficients of individual background characteristics in the baseline model (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo maximum likelihood method

Parameters	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
(Coefficient on)	(1)	(2)	(3)
Constant	-0.68177	-0.03845	-0.87633
	(3.69140)	(3.53725)	(3.58852)
7th Grade	0.47755	0.46843	0.46602
	(0.08802)	(0.08763)	(0.08605)
8th Grade	0.94492	0.93562	0.93465
	(0.08213)	(0.08088)	(0.08058)
9th Grade	0.98692	0.97299	0.99142
	(0.08437)	(0.08285)	(0.08340)
10th Grade	1.07936	1.06156	1.07325
	(0.08838)	(0.08718)	(0.08789)
11th Grade	1.07958	1.05034	1.06054
	(0.09053)	(0.08826)	(0.08822)
12th Grade	1.15153	1.13572	1.15289
	(0.09397)	(0.09157)	(0.09232)
Asian	-0.54672	-0.54873	-0.54734
	(0.13176)	(0.12929)	(0.12993)
Black	-0.80229	-0.80005	-0.79480
	(0.07247)	(0.07182)	(0.07200)
Hispanic	-0.39855	-0.41438	-0.42327
	(0.07055)	(0.06743)	(0.06971)
White Male	-0.12576	-0.14919	-0.13157
	(0.06984)	(0.06733)	(0.06944)
Asian Male	0.47027	0.43158	0.44862
	(0.17744)	(0.17336)	(0.17580)
Black Male	0.24751	0.20722	0.27635
	(0.11557)	(0.11408)	(0.11248)
Hispanic Male	0.22300	0.18893	0.25128
	(0.10729)	(0.10414)	(0.10617)
Weekly Income	0.02878	0.02903	0.02911
	(0.00231)	(0.00224)	(0.00226)
Working Dummy	0.31798	0.32408	0.32637
	(0.04204)	(0.04144)	(0.04216)
Smokers in Home	0.96354	0.96556	0.96087
	(0.03308)	(0.03181)	(0.03222)
See Actors Smoking in TV	0.05762	-	0.07542
	(0.04963)	-	(0.04378)
See Actors Smoking in Movie	0.05875	0.06059	-
	(0.05656)	(0.04968)	-
School Program 1	-0.19903	-0.20304	-0.18129
	(0.04626)	(0.04410)	(0.04499)
School Program 2	-0.09634	-0.10557	-0.11811
	(0.04450)	(0.03855)	(0.03880)
School Program 3	0.14327	0.16336	0.14783
	(0.04733)	(0.04570)	(0.04590)
School Program 4	-0.01018	-	-
	(0.04179)	-	-
Cigarette Tax	-0.23608	-0.21771	-0.20983
	(0.09212)	(0.08868)	(0.08925)

Asymptotic standard errors are in parentheses.

TABLE 5. The estimated peer effects in the baseline model (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo maximum likelihood method

Peer Effects	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
	(1)	(2)	(3)
ρ_{MM}	1.70269	1.66938	1.66987
	(0.20653)	(0.19879)	(0.20741)
ρ_{FF}	1.65828	1.73793	1.62800
	(0.20407)	(0.19796)	(0.20646)
ρ_{MF}	0.72750	0.74057	0.76848
	(0.16668)	(0.16465)	(0.17157)

Asymptotic standard errors are in parentheses.

TABLE 6. Distribution of Schools with Respect to Racial Composition: Number of Schools in Each Category

	Percentage of Each Race Group									
	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
White	48	29	16	15	17	22	29	25	38	79
Black	180	41	27	21	15	6	4	6	8	10
Hispanic	187	32	28	16	12	6	5	6	16	10

TABLE 7. Estimated raced-based peer effects (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo maximum likelihood method

Peer Effects	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
	(1)	(2)	(3)
ρ_{WW}	1.55833	1.58750	1.53104
	(0.20126)	(0.19402)	(0.20347)
ρ_{BB}	0.77466	0.79752	0.72025
	(0.36861)	(0.36423)	(0.36913)
ρ_{HH}	1.44520	1.39753	1.45932
	(0.36970)	(0.35295)	(0.35689)
ρ_{WB}	0.01568	0.03009	-0.00037
	(0.25102)	(0.25006)	(0.25667)
ρ_{WH}	0.71963	0.75615	0.71113
	(0.27574)	(0.26759)	(0.28123)
ρ_{BH}	0.99595	0.94710	0.88158
	(0.33315)	(0.33104)	(0.34128)

Asymptotic standard errors are in parentheses.

TABLE 8. The estimated coefficients of individual background characteristics in the model with fixed effects (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo conditional maximum likelihood method

Parameters	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
(Coefficient on)	(1)	(2)	(3)
Constant	-	-	-
7th Grade	0.52662	0.52286	0.50752
	(0.09180)	(0.09025)	(0.09007)
8th Grade	1.01922	1.00826	0.99930
	(0.08510)	(0.08356)	(0.08502)
9th Grade	1.01244	0.98488	1.00323
	(0.09690)	(0.09490)	(0.09374)
10th Grade	1.19693	1.17840	1.17205
	(0.10006)	(0.09764)	(0.09762)
11th Grade	1.21891	1.19134	1.18839
	(0.10136)	(0.10007)	(0.09815)
12th Grade	1.31568	1.29056	1.29675
	(0.10745)	(0.10483)	(0.10376)
Asian	-0.64295	-0.63722	-0.65163
	(0.14644)	(0.14265)	(0.14941)
Black	-0.90837	-0.90505	-0.89120
	(0.08222)	(0.07740)	(0.07982)
Hispanic	-0.34677	-0.35483	-0.37839
	(0.07724)	(0.07665)	(0.07645)
White Male	-0.13693	-0.16880	-0.12908
	(0.07756)	(0.07588)	(0.07682)
Asian Male	0.58284	0.52515	0.58495
	(0.19558)	(0.18740)	(0.19418)
Black Male	0.25413	0.21341	0.30129
	(0.11848)	(0.11556)	(0.11746)
Hispanic Male	0.21696	0.16202	0.25584
	(0.11501)	(0.11355)	(0.11223)
Weekly Income	0.03073	0.03055	0.03084
	(0.00242)	(0.00242)	(0.00241)
Working Dummy	0.30452	0.31242	0.31701
	(0.04563)	(0.04442)	(0.04414)
Smokers in Home	0.96465	0.96668	0.95985
	(0.03522)	(0.03420)	(0.03567)
See Actors Smoking in TV	0.07796	-	0.09351
	(0.05369)		(0.04759)
See Actors Smoking in Movie	0.05611	0.05994	-
	(0.05931)	(0.05173)	
School Program 1	-0.20497	-0.20845	-0.18925
	(0.04906)	(0.04788)	(0.04754)
School Program 2	-0.07463	-0.09275	-0.10200
	(0.04742)	(0.04181)	(0.04177)
School Program 3	0.15234	0.15958	0.14387
	(0.04949)	(0.04791)	(0.04903)
School Program 4	-0.03682	-	-
	(0.04494)		
Cigarette Tax	-	-	-

Asymptotic standard errors are in parentheses.

TABLE 9. The estimated coefficients of peer effects in the model with fixed effects (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo conditional maximum likelihood method

Peer Effects	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
	(1)	(2)	(3)
ρ_{MM}	1.52090	1.47377	1.54619
	(0.24665)	(0.25002)	(0.25164)
ρ_{FF}	1.48030	1.60347	1.46383
	(0.24326)	(0.23562)	(0.24089)
ρ_{MF}	0.52547	0.56642	0.58560
	(0.20739)	(0.20826)	(0.20620)

Asymptotic standard errors are in parentheses.

TABLE 10. The estimated coefficients of race-based peer effects in the model with fixed effects (smoking choices among students in grades six to 12): The estimates are calculated by using the Monte Carlo conditional maximum likelihood method

Peer Effects	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
	(1)	(2)	(3)
ρ_{WW}	1.49926	1.50204	1.47278
	(0.31725)	(0.30838)	(0.31973)
ρ_{BB}	0.78170	0.73457	0.63131
	(0.53012)	(0.52385)	(0.52662)
ρ_{HH}	0.51174	0.51186	0.58596
	(0.55511)	(0.54107)	(0.53320)
ρ_{WB}	-0.18108	-0.24148	-0.25722
	(0.37773)	(0.37937)	(0.38538)
ρ_{WH}	0.29493	0.36871	0.30975
	(0.43672)	(0.43398)	(0.42431)
ρ_{BH}	-0.44739	-0.42444	-0.48892
	(0.52332)	(0.51826)	(0.53077)

Asymptotic standard errors are in parentheses.

TABLE 11. Simulated Smoking Prevalence Following Tax Increases

	Baseline	Hypothetical tax increases (cent)					
	Simulation	20	40	60	80	100	120
Predicted smoking rates	18.948	18.056	17.223	16.407	15.65d	14.922	14.213
Estimated tax elasticities		0.210	0.203	0.199	0.194	0.190	0.199

All values are computed from the average over 1000 experiments.

Tax elasticities are based on the average tax of 89.26 cent in sample.

TABLE 12. Decomposition of Tax Effects on Smoking Prevalence

	Hypothetical tax increases (cent)					
Type of effects	20	40	60	80	100	120
Total effects	-0.892	-1.725	-2.540	-3.294	-4.026	-4.735
Direct effects	-0.541	-1.070	-1.600	-2.121	-2.614	-3.106
Social multipliers	1.649	1.612	1.588	1.553	1.540	1.525

All values are computed from the average over 1000 experiments.

Social multiplier is defined by the ratio of total effect to direct effect.

TABLE 13. Simulated Smoking Prevalence Following Implementation of School-based Smoking Program 1

	Baseline	Hypothetical School Program increases (percent)				
	Simulation	10	20	30	40	50
Predicted smoking rates	18.640	18.078	17.924	17.522	17.172	16.813
Estimated elasticities		0.082	0.052	0.054	0.053	0.053

All values are computed from the average over 1000 experiments.

Elasticities are based on the average school program of 27.04 percent in sample.

APPENDIX D. APPENDIX TABLES FOLLOW

TABLE 14. the estimated coefficients of county characteristics in the model with fixed effects (smoking choices among students of 6-12th grades): The estimates are calculated by the Monte Carlo maximum likelihood methods

Parameters	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications		
(Coefficient on)	(1)	(2)	(3)
Persons 0-18	-0.00531	-0.00856	-0.00226
	(0.02280)	(0.02216)	(0.02218)
Persons 65+	0.00251	0.00311	0.00332
	(0.00790)	(0.00768)	(0.00787)
Asian persons	0.00627	0.00618	0.00655
	(0.00530)	(0.00521)	(0.00525)
Black persons	0.00327	0.00290	0.00356
	(0.00344)	(0.00334)	(0.00341)
Hispanic persons	0.00491	0.00575	0.00531
	(0.00223)	(0.00221)	(0.00222)
Single-mother families	0.00550	0.01185	0.00581
	(0.01550)	(0.01373)	(0.01518)
One-person families	-0.02210	-0.02406	-0.02212
	(0.00882)	(0.00860)	(0.00873)
Persons with grandchildren	-0.04468	-0.04420	-0.04561
	(0.02475)	(0.02395)	(0.02453)
Persons who divorced	-0.01008	-0.01058	-0.00903
	(0.01431)	(0.01412)	(0.01425)
Unemployed workers	0.04817	0.04412	0.04669
	(0.01418)	(0.01388)	(0.01400)
Workers with management job	-0.01377	-0.01315	-0.01367
	(0.01210)	(0.01164)	(0.01160)
Workers with professional job	-0.01541	-0.01704	-0.01614
	(0.01209)	(0.01188)	(0.01180)
Workers with construction job	0.01764	0.01721	0.01608
	(0.01191)	(0.01140)	(0.01161)
Workers with farming job	-0.01132	-0.01089	-0.01188
	(0.01191)	(0.01141)	(0.01178)
Workers with sales and office job	0.00966	0.00675	0.00829
	(0.00844)	(0.00816)	(0.00814)
Workers with service job	0.00188	0.00190	0.00129
	(0.00983)	(0.00931)	(0.00957)
Persons below poverty level	-0.00474	-0.00575	-0.00318
	(0.01062)	(0.01013)	(0.01028)
Persons graduated high school	-0.01024	-0.01058	-0.00911
	(0.00596)	(0.00580)	(0.00595)
Persons graduated college	0.01706	0.01802	0.01838
	(0.00769)	(0.00743)	(0.00746)

(continued on next page)

Asymptotic standard errors are in parentheses.

Parameters (Coefficient on)	$\mathbf{x}_i\boldsymbol{\beta}$ Specifications (1)	(2)	(3)
(continued from previous page)			
Persons enrolled in public schools	0.00192	0.00112	0.00196
	(0.00391)	(0.00376)	(0.00391)
Persons enrolled in high schools	-0.02973	-0.02786	-0.03177
	(0.03017)	(0.02911)	(0.02927)
Persons enrolled in colleges	-0.00853	-0.00784	-0.00872
	(0.01254)	(0.01204)	(0.01206)
Persons enrolled in graduate schools	0.07041	0.06842	0.06650
	(0.05416)	(0.05306)	(0.05368)
Persons in armed forces	0.02012	0.01924	0.01878
	(0.01215)	(0.01184)	(0.01196)
Persons in veteran status	0.01126	0.01234	0.01312
	(0.01091)	(0.01049)	(0.01089)
Persons imigrated from foreign countries	-0.03224	-0.03264	-0.03157
	(0.01757)	(0.01680)	(0.01691)
Persons moved in last 5 years	-0.01254	-0.01212	-0.01241
	(0.00362)	(0.00346)	(0.00350)
Median household income	0.35930	0.31293	0.36120
	(0.38911)	(0.37372)	(0.37826)
Median housing value	-0.31850	-0.30472	-0.31418
	(0.09055)	(0.08617)	(0.08860)
Vacant housing units	-0.00103	-0.00402	-0.00249
	(0.00264)	(0.00129)	(0.00271)
Private vehicles occupacy	-0.00305	-0.00181	-0.00102
	(0.00272)	(0.00258)	(0.00260)
Urban area rate	-0.00385	-0.00315	-0.00384
	(0.00136)	(0.00257)	(0.00128)
Crime index	-0.01649	-0.01675	-0.01898
	(0.01373)	(0.01334)	(0.01386)
Juvenile crime	0.00732	0.00668	0.00880
	(0.01442)	(0.01397)	(0.01427)
Juveniles crime (drug abuse)	0.18146	0.19590	0.18956
	(0.16581)	(0.16016)	(0.16496)

Asymptotic standard errors are in parentheses.