Review of Economic Studies (2007) **00**, 1–41 0034-6527/07/00000001$02.00

Measuring Peer Effects on Youth Smoking Behavior

RYO NAKAJIMA

University of Tsukuba, Graduate School of Humanities and Social Sciences
First version received July 2004; final version accepted November 2006 (*Eds.*)

This paper examines the role of peer effects in teenagers' smoking behavior in the United States. I present a random utility model that incorporates complementarity between individual and peer smoking. A Markov process model of smoking interactions between individuals is presented. I estimate the structural parameters of the model using a steady state distribution that is determined by the Markov process. The empirical results strongly support the presence of positive peer effects. 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.

1. INTRODUCTION

The prevalence of youth smoking is a major public health concern in the United States. 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, in the 1990s, the smoking rate increased by about a half among 8th and 10th graders and by nearly one-third among 12th graders. Although smoking has declined since the late 1990s and into the 2000s, nearly a quarter of youths are smokers by the time they complete high school (Johnston et al. 2004). Because smoking at early age leads to long-term health consequences in later life, preventing smoking among young people is critical to ending the epidemic of diseases related to tobacco use in the United States.

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. For example, Chaloupka and Pacula (1999) find that the price elasticity is -0.93 for male high-school students while the price elasticity is -0.60 for female high-school students, based on micro-data from the 1992-1994 Monitoring the Future Survey. They also find that black male students are the most responsive to price; in this case the estimated price elasticity is -1.65. Gruber and Zinman (2000), based on the 1991-1997 Monitoring the Future Survey data, find that the price elasticity is -0.35 for white high-school students while the price elasticity is -2.32 for black high-school students.[1] However, these results raise the question of why gender and race so markedly affect the elasticity of demand for smoking.

One explanation is that the underlying cause of this 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 not only have a direct effect on consumption but also have an indirect effect—changes in the consumption level

1. See also *Report of the Surgeon General* (1998) .

of the peer group as a whole affect demand among individual group members. Thus, differences in the intensity of such interactions could account for differences in the price elasticity between groups.

In recent years, economists and other social scientists have devoted much effort to studying peer interactions in smoking behavior among young people (e.g., Norton et al. 1998; Gaviria and Raphael 2001; Powell et al. 2005). Peer effects have been of interest because they imply an externality that can lead to large differences in smoking behavior through social-multiplier effects. Given the presence of a strong peer effect, government interventions to prevent young people from starting to smoke— mandated tobacco education in schools, a complete ban on smoking by anyone on school grounds, restrictions on tobacco advertising, and prohibitions on the sale of tobacco products to minors — might be facilitated further.

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 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 endogenous aspect of peer choice causes simultaneous equations bias. Peer choice, which enters a person's utility function reciprocally, is simultaneously affected by that person's choice. 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.

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 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.

In order to address the difficulty described above, I adopt the following empirical strategies for estimating a social interaction model that incorporates peer effects.

First, I distinguish between endogeneity and simultaneity in decision making. Endogeneity of choice follows from simultaneous decision making, but simultaneity is not the only cause of choices being endogenous. I assume that youth smoking decisions occur *sequentially* rather than simultaneously. That is, whereas the simultaneity of choices is not necessarily an essential feature of the social interaction model, the reciprocity of choices is. Choices are considered reciprocal if the direction of influence in social interactions is *two-way.* [2] As reciprocal interactions are repeated, smoking profiles of a group of persons

2. Reciprocity emphasizes the bidirectional influence of social interactions. To make the difference in unidirectional or bidirectional influence of social interactions clear, consider two similar but distinct

evolve together over time. Feedback though peer-to-peer interactions generates a positive correlation between the smoking choices of individuals in the same peer group. Within the sequential framework, I argue that all outcomes are endogenous variables that are determined by the model itself. The joint distribution of outcomes is considered in terms of the likelihood function.

Second, I include county fixed effects to account for unobserved common factors. A model with fixed effects allows the independent identification of endogenous peer effects and exogenous correlated effects, under the assumption that all omitted variables vary only at the county level. I exploit the cluster nature of the NYTS data set to control for unobserved characteristics common to schools in the same county. Endogenous peer effects are identified by using variations in the proportion of smokers between schools within a county. Fixed-effects approaches have been used in recent empirical studies of peer effects (e.g., Bertrand et al. 2000; Weinberg et al. 2004; Arcidiacono and Nicholson 2005).

The main methodological tool of this paper is a variant of the framework of best-response dynamics, which is similar to Blume (1993). Under the assumption that peer interactions occur frequently, a best-response dynamic model can determine the evolution of smoking profiles. Every person, given the opportunity to review his or her smoking status, updates the smoking choice to maximize his or her current utility while treating other persons' smoking as exogenous. Idiosyncratic taste shocks lead to stochastic transitions from one state to another in smoking profiles. I show that the dynamic interaction process follows a Markov chain on finite spaces of the smoking profile.

In the model of dynamic interaction process, people are assumed to make the best response to the observed smoking or nonsmoking actions of others, and they do not attempt to anticipate the actions of others in the future. Thus, this paper assigns a very limited view to human rationality about the ability to forecast others' behavior. The empirical method developed in this paper may not be applied to the study of social interactions where the expectation of others' behavior plays an important role.[3] Yet, I believe, the model in this paper contributes to a plausible visualization of interactive smoking decisions among teenagers. The strategy of this paper is to demonstrate how a simple form of interactive behavior leads to a predictable pattern of smoking behavior.

The result of this paper establishes that the dynamic interaction process converges to a unique steady-state distribution as the number of periods becomes large, and that it is independent of the initial state from which the dynamic process starts. The steady-state distribution, which is defined over all possible states of the smoking profile, provides a precise prediction of the smoking state most likely to prevail in the long run.

I estimate the structural parameters of the social interaction model by using the steady-state distribution of the interaction process. The steady-state distribution is assumed to represent a cross-sectional state 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 by maximum likelihood.

social interaction effects: peer effect and role-model effect. Peer effect is considered as reciprocal, as it refers to a two-sided influence in imitation, while role-model effect is considered as nonreciprocal, as it refers to a one-sided influence in imitation.

3. For example, consider the study of social interactions on fertility choice. In this case, individuals' expectations regarding each other would play an important role in a family planning problem. See Durlauf and Walker (2001) for example.

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 analytically intractable, it is approximated by using a simulation method. Simulated samples are drawn from the Markov chain of the dynamic interaction process described above. This technique was developed by Geyer and Thompson (1992). The second set of maximum likelihood estimates incorporates fixed effects into the social interaction 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 of students taking the same compulsory courses in each school, 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 (in as much as they are captured by county effects).

A variety of additional specifications are also examined to determine the robustness of the findings. I consider less restrictive assumptions on types of peers and introduce a new characteristic to define peer type. Nevertheless, positive and significant peer effects are found. Furthermore, separate models are estimated for middle-school and high-school students to address heterogeneity in smoking behavior due to "addictive stock". The results show that high-school students are less susceptible to peer pressure than are middle-school students. This is consistent with a priori expectations that addiction might dampen the magnitude of peer effects. Nevertheless, sizable positive peer effects are found not only for middle-school students but also for high-school students.

The paper also examines the expected response of youth smoking behavior to changes in hypothetical smoking policies. The simulation results strongly indicate that cigarette excise tax is an important policy tool for discouraging youths from smoking. Policy experiments based on the estimated social models show that a 10 percent increase in the tax on cigarettes could reduce the youth smoking rate by about 2 percent. Furthermore, tax increases can explain about a third of the decline in smoking among both middle-school and high-school students in the early 2000s. A multiplier effect is also found. The impact of a tax on youth smoking increases by a factor of more than 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 6, I discuss the robustness and policy implications of these findings. Section 7 concludes the paper. Proofs and derivations of some ancillary results are presented in the appendixes.

2. BEHAVIORAL MODEL

2.1. *Best Response Revision*

I construct a simple interaction model based on the random-utility framework of binary choice. The critical feature is that the chance that a young person will temporarily take up smoking increases with the fraction of smokers in his or her peer group. Such temporary smoking might be considered as pure experimentation for young persons, so that a puff of a cigarette might be driven by strong peer pressure. Although several alternative models are possible for peer interaction models,[4] I adopt a simple discrete-choice framework by using a parameterization proposed by Brock and Durlauf (2001).

Suppose that there are N persons. Persons are indexed by $i \in I \equiv (1, \cdots, N)$. So, the set I denotes a *peer group* of N persons. Suppose that person $i \in I$ is deciding whether or not to smoke cigarettes. Let $y_i \in \{-1, +1\}$ denote the smoking status of person i. I assume that smoking status is *binary* such that $y_i = +1$ if the person is smoking, and $y_i = -1$ otherwise.

A *smoking profile* $\mathbf{y} = (y_1, \cdots, y_N)$ is a vector of the smoking status of all N persons. Let $\Omega = \{-1, +1\}^N$ denote all possible states of the smoking profile. The number of different states of Ω is given by $|\Omega| = 2^N$.

Persons get satisfaction, or utility, from smoking cigarettes. Let y_i^* denote the *latent utility* from smoking for person i. I assume that utility is given by the following function, which is linear in the parameters:

$$y_i^* = b_i(\mathbf{x}_i) + \textstyle\sum_{j \neq i} \rho_{ij} y_j + \epsilon_i. \tag{2.1}$$

The first 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. Let $f(\epsilon)$ and $F(\epsilon)$ be the density and distribution functions of the shock ϵ, respectively. 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.[5] The second utility component involves the social capital of the smoking behavior of other persons. This component is given by $(\sum \rho_{ij} y_j)$. The parameter ρ_{ij} measures conformity; i.e., the degree to which person i behaves like person j. In short, the parameter ρ_{ij} represents the *peer effect* between person i and j.

I assume that the peer effect does not depend on any individual characteristics other than *types*. In other words, the peer effect ρ_{ij} is assumed to be constant between person i and person j, who belong to the same type. The assumption will be discussed in the later section, which is concerned with coping with empirical specification of the model.

Decisions are made to maximize latent utility. While smoking utility is given by equation (2.1), nonsmoking utility is normalized to zero. An individual with positive latent utility chooses to smoke. Let $y_{-i} \equiv \{y_j, j \in I \backslash \{i\}\}$ be a smoking profile comprising the smoking status of the $(N - 1)$ persons other than person i in the peer group I.

4. See Glaeser and Scheinkman (2000) for various social interaction models.
5. For example, the random variable ϵ_i could represents exposure to psychological stress that may lead a person to start smoking.

Assuming that the stochastic errors ϵ_is are independent and identically distributed across persons, the probability that person i smokes is $(y_i = +1)$, conditional on $\mathbf{x}_i$ and y_{-i}, is given by

$$\pi_i(y_i = +1|y_{-i}, \mathbf{x}_i) = \int_D f(\epsilon_i) d\epsilon_i, \tag{2.2}$$

where $D = \{\epsilon_i \in \mathbb{R} | y_i^* > 0\}$ denotes the area under which the integral is taken. Equation (2.2) represents a *stochastic best-response rule.*

It is very difficult to estimate Equation (2.2) directly by using microeconomic data. The problem is that the information about the other person's smoking status y_{-i} is that person i's conditions at the time of his smoking decision are hardly available from data. I will discuss an alternative estimation strategy in detail in the next section.

2.2. *A Dynamic Interaction 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 but are 'on-again-off-again' events. This assumption is supported in many studies in developmental psychology (e.g., Flay et al. 1983) in which it is argued that most teenagers experiment with smoking sufficiently repeatedly to acquire the smoking habit.

To be more precise, I consider a stochastic process in which each person's smoking profile develops in discrete steps. Therefore, it is convenient to use discrete time, $t = 0, 1, 2, \cdots \in \mathbb{Z}$. In what follows, I use $y_i(t)$ and $\mathbf{x}_i(t)$ to denote the smoking status of person i at time t and a vector of characteristics on that person, respectively.

The specification of the stochastic process relies heavily on local interaction models of learning and adaptive behavior from game theory (e.g., Blume 1993). The key features of these models 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.[6] Let $d(t) \in I$ be the person who makes a decision at time t. In a sufficiently small interval of time, it is *unlikely* that two or more persons will make decisions simultaneously. Hence, it is reasonable to assume that the decisions occur *sequentially*, so that only one person $d(t)$ is selected out of the peer group I to make a decision at each moment $t = 0, 1, 2, \cdots \in \mathbb{Z}$. Thus, decisions are given by a sequence $(d(0), d(1), d(2), \cdots)$.

In adaptive behavior, a person makes a decision by considering the current, not expected future, rewards of each choice. Let $y_i^*(t)$ be the latent utility derived from smoking by person i at time t. Then, analogously to the way that latent utility is represented by equation (2.1) from the static model, I assume that

$$y_i^*(t) = b_i(\mathbf{x}_i(t)) + \textstyle\sum_{j \neq i} \rho_{ij} y_j(t-1) + \epsilon_i(t). \tag{2.3}$$

In other words, person i at time t chooses between $y_i(t) = +1$ if $y_i^*(t) \geq 0$ and $y_i(t) = -1$ if $y_i^*(t) < 0$, while treating other persons' smoking, $y_j(t-1)$, as exogenous. Each person is given an opportunity to revise their choice responding to the lagged decisions of others in the peer group.

I would also like to emphasize the scope of social interactions that occur between people who imitate the smoking behavior of others. The model's local interaction framework postulates that in every period, only one person is making a smoking decision

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

after observing the lagged choices of others in the *same* peer group. As described later in the empirical part of the paper, I assume that the structure of the economy is such that the entire population is divided into a number of mutually exclusive peer groups of relatively small size. It thus follows that each process of smoking interaction is locally independent across peer groups.

Note that I adopt a relatively simple specification for studying youth smoking decisions; i.e., one that excludes cumulative past smoking and the 'addiction stock' from equation (2.3). I assume that each person is not yet addicted when making a decision about temporary smoking. As the rational addiction model (e.g., Becker and Murphy 1988) implies, the stock of addiction plays an important role in adult smoking. However, 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. Hence, it might be reasonable to assume that youth smoking is not influenced by expected future smoking. Second, as explained in section 4, data show that young people have low levels of past smoking.[7] Note also that, although addiction to smoking is not fully taken into account, differences in smoking behavior due to different levels of addictive stocks are examined in a later section.

The specific way in which I model smoking interactions is through a discrete-time stochastic process in which each person updates his or her smoking choice sequentially over time. Let $\mathbf{y}(t) \equiv (y_1(t), \cdots, y_N(t)) \in \Omega$ denote a smoking profile at time t. Then, a sequence $[\mathbf{y}(0), \mathbf{y}(1), \mathbf{y}(2), \cdots]$ describes the evolution of smoking profiles over time. The transition from one state to another is specified as follows. Suppose that a smoking profile is $\mathbf{y}(t) = \boldsymbol{\omega} = (\omega_1, \cdots, \omega_N)$ at time t. Then a new smoking profile in period $t+1$ evolves from the smoking profile in period t according to the following transition. Let $\mathbf{x}(t) \equiv (\mathbf{x}_1(t), \cdots, \mathbf{x}_N(t))$ be the collection of background characteristics for N persons at time t. For each smoking status $\nu \in \{-1, +1\}$,

$$y_i(t+1) = \begin{cases} \nu & \text{if } i = d(t), \\ \omega_i & \text{if } i \neq d(t), \end{cases} \tag{2.4}$$

with $\text{Prob}(y_i(t+1) = \nu | \mathbf{x}(t)) = \pi_i(y_i = \nu | y_{-i} = \omega_{-i}, \mathbf{x}_i(t))$ for $i = d(t)$. Recall that the assumption of sequential decisions allows person $d_t \in I$ to review his or her smoking status in period t. The transition rule states that the smoking status of person $d(t)$ is updated according to the conditional probability represented by the best-response rule and given by equation (2.2). However, the smoking status of persons other than $d(t)$ remains unchanged.

I refer to the stochastic process $[\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(t))]$ described above as an *interaction process*. It is simple to check that the transition probability at time $t+1$ is independent of its history before time t. Thus, the interaction process follows a *Markov chain* on a finite state space of Ω. Markov chains 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).

I make the following three assumptions about the interaction process.

7. Note however that the model assumes that the stock of addictive capital has a negligible effect on youth smoking. Empirical analysis of rational addiction processes with peer effects warrants further research. For example, Bisin et al. (2006) study the rational expectations equilibria of a model with peer interactions and incomplete information.

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

Assumption 2. The vector of characteristics is time invariant: $\mathbf{x}(t) = \mathbf{x}$ for any period $t \in \mathbb{Z}$.

Assumption 3. $\text{Prob}(d(t) = i) > 0$ for any person $i \in I$ and for any period $t \in \mathbb{Z}$.

The first assumption concerns the error distribution. The assumption of the logistic distribution is standard in the literature (see Brock and Durlauf 2001). The second assumption does not necessarily require that the characteristics be constant over time.[8] The third assumption requires that every person be able to make a decision in each time period.

It is straightforward to show that the interaction process $[\mathbf{y}(0), \mathbf{y}(1), \cdots, \mathbf{y}(t))]$ is an aperiodic and irreducible Markov chain. The standard result shows that if the Markov chain is aperiodic and irreducible, it is asymptotically convergent to the unique *steady-state distribution.* I present the asymptotic distribution subsequently.

2.3. *Steady-State Distribution*

The following result specifies the steady-state distribution of the interaction process.

Theorem. Let assumptions 1–3 hold. (i) The interaction process has a unique steady-state distribution P^* such that, for any $\boldsymbol{\omega}$ and $\boldsymbol{\omega}(0) \in \Omega$,

$$\lim_{t\to\infty} \text{Prob}(\mathbf{y}(t) = \boldsymbol{\omega}|\mathbf{y}(0) = \boldsymbol{\omega}(0), \mathbf{x}(0) = \mathbf{x}) = P^*(\mathbf{y} = \boldsymbol{\omega}|\mathbf{x}). \tag{2.5}$$

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

$$P^*(\mathbf{y} = \boldsymbol{\omega}|\mathbf{x}) = \exp Q(\boldsymbol{\omega}|\mathbf{x}) \Big/ \sum_{\boldsymbol{\eta}\in\Omega} \exp Q(\boldsymbol{\eta}|\mathbf{x}), \tag{2.6}$$

where

$$Q(\boldsymbol{\omega}|\mathbf{x}) = \frac{1}{2}\sum_i \omega_i b_i(\mathbf{x}_i) + \frac{1}{4}\sum_i \sum_j \rho_{ij}\omega_i\omega_j, \tag{2.7}$$

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

The proof of the theorem is presented in the Appendix. It is an application of well-known results concerning the convergence of Markov chains.

To specify the probability structure for the steady state distribution, the assumption that only one person is selected out of one peer group at each moment, or that what Blume (2003) calls the "one person at a time" formalism, is crucial. As Blume (1993, 2003) argued, the stochastic dynamic process that allows for simultaneous decision making *inside* each peer group would yield a different equilibrium state from the one presented in this paper.[9]

8. However, it does require the background characteristics to remain invariant, at least over the time-scale for decision making. The assumption is reasonable if decisions about temporary smoking change more frequently (e.g., day-to-day) than background characteristics change (e.g., year-to-year).

9. The "one person at a time" assumption of decision making has been widely used in economic analysis of social (strategic) interactions. For example, See Young (1993);Glaeser and Scheinkman (2000);Blume and Durlauf (2003); Blume (2003).

The feature of "one person at a time" assumption distinguishes the discrete-choice model of this paper from other similar models of social interactions. For example, Brock and Durlauf (2001) assumes that each person makes his or her decision simultaneously by responding to the common expectation of other choices. Idiosyncratic shocks are assumed to be private information, so that each person is uncertain what decisions are being made by others at the same time as his or her own. Thus, their model is characterized by incomplete information. On the other hand, the sequential decisions model considered in this paper assumes that individuals do not attempt to anticipate the actions of others. Rather, each individual is assumed to take the best response to the observed choices of others with a lag. [10]

It is important to note that the simultaneous decisions model of social interactions yields multiple equilibria. A model with multiple equilibria has neither a unique reduced form probability nor a well-defined joint probability distribution. In this case, point estimation by maximum likelihood is not generally feasible. On the other hand, the interaction process regarding the sequential decisions does not lead to 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 steady-state distribution allows each state to be revisited. Thus, the stochastic process does not converge to a few distinct equilibria.

One final note concerns the possibility of contemporaneous decision making by multiple persons in the economy as a whole. The assumption of sequentiality, regarding the revision opportunity, may be considered too artificial for the *global* interaction framework in which every person interacts with all others in the economy. It implies that the revision opportunities of a group increase with the size of the group. Therefore, the interval between the revision opportunities must be taken to be perhaps unrealistically small for the global interaction economy of a large population. However, the *local* interaction nature of the model allows the interaction process of smoking decisions, presented by the best-response rule of equation (2.3), to be independent across peer groups. It thus follows that multiple numbers of people who are members of different peer groups might be able to make their smoking decisions contemporaneously in the *whole* economy, while only one person who is selected out of each peer group is making a smoking decision sequentially in *part* of the economy. Moreover, no matter how large the entire economy becomes, the size of each peer group is fixed, with the number of peer groups increasing. Hence, the assumption of sequential decision making seems no less realistic than the alternative.

3. EMPIRICAL STRATEGY

In this section, I discuss the estimation of the behavior model described above. The basic idea is that an observed smoking profile can be taken as a *realization* of the steady-state distribution of the interaction process.

10. Other simultaneous decisions models of social interactions are also proposed. For example, Heckman (1978) , Bresnahan and Reiss (1990) and Tamer (2003) assume that decision making also takes place simultaneously but that idiosyncratic shocks are *common knowledge* rather than private information. In these models, each person observes the realizations of others' idiosyncratic random shocks and has complete information about others' choices.

3.1. *The Scope for Peer Group Interaction*

For estimation, one needs to define each person's peer group. In this paper, I assume that a person's *school* cohort represents a well-defined peer group. Moreover, I assume that a person interacts daily with others in the same school. In other words, smoking interactions occur *within*, not between, schools. These assumptions seem reasonable given the absence of information about the structure of a person's peer group. Evidence from the sociology and social psychology literature indicates that the majority of middle-school and high-school students choose as their peers fellow students from the same schools.[11]

I introduce several pieces of notation. Suppose that there are S different schools observed in the data set. Schools are indexed by $s \in \{1, 2, \cdots, S\}$. Every person attends one of the finite number of nonoverlapping schools. Let $I_s \subset I$ denote a set of persons at school s and $N_s \equiv |I_s|$. For expositional simplicity, for now, I assume that schools are the same size: $N_s = N$ for all $s = 1, \cdots, S$. Let $\mathbf{y}_s \equiv \{y_i, i \in I_s\}$ be a smoking profile at school s, which is a vector of smoking status for N persons. Let $\mathbf{x}_s \equiv \{\mathbf{x}_i, i \in I_s\}$ be an $N \times K$ matrix of individual characteristics of N persons at school s. In what follows, I use Ω_s to denote all possible states of $\mathbf{y}_s$. Then, $\mathbf{y}_s \in \Omega_s$.

3.2. *The Likelihood Function*

The objective is to estimate the structural parameters $(b_i, \rho_{ij} : i, j \in I)$ of the latent-utility model (equation (2.1) and (2.3)). 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. This type of linear specification is quite standard in the literature on discrete-choice theory.

Assumption 4. The perceived benefits of 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.

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 $g \in \{1, 2, \cdots, G\}$. Then, one can state the following.

Assumption 5. 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 $\rho_{ij} = \rho_{gg'}/N$.

As an example person types, I consider *gender*. The most consistent finding of the literature on peer groups (e.g., Shrum et al. 1988; 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.

Another example type is *race*. Many studies in sociology provide evidence that peer groups are formed along racial and ethnic lines. Assuming that *type*

11. Shrum et al. (1988), based on studies of friendship structures among 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.

g is defined by race $\{W, B, H\}$, the following peer effects are considered: $\boldsymbol{\rho} = (\rho_{WW}, \rho_{BB}, \rho_{HH})$ and $(\rho_{WB}, \rho_{BW}, \rho_{BH}, \rho_{HB}, \rho_{HW}, \rho_{WH})$, 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}. [12]

The behavioral model described in the last section can be represented in terms of the parameterization described above. For notational convenience, let the subscript g represent type. Hence, I let y_{ig} be the smoking status of person i of type g, and let y^*_{ig} be latent utility. Then, equation (2.3) can be written in recursive form as follows:

$$y^*_{ig}(t) = \alpha + \mathbf{x}_i(t)\boldsymbol{\beta} + \textstyle\sum_{g'} \rho_{gg'} \left(\frac{1}{N} \sum_j y_{jg'}(t-1) \right) + \epsilon_i(t). \tag{3.8}$$

This implies that a person updates his or her smoking choice by responding to the average choice of each subgroup g observed in the previous period. The smoking influences are transmitted through peer effects with a lag. For now, unobservable errors are assumed to be independent across persons i. The possibility of correlated unobservables is addressed in a later section.

At first glance, it might seem straightforward to estimate equation (3.8) by using microeconomic data. In practice, however, it is difficult (and can be impossible) to estimate the equation by using the available data. The problem is that it is not always possible to determine what individuals know at the time of their *actual* decisions. Cross-section data has no information on whose choice precedes whose. It is impossible from this "snapshot" of choices to observe the choices of the reference group to which persons respond. Panel data would not solve this problem. To estimate the sequential choice model presented above, one needs to know the precise order in which decisions are made.[13] However, the sampling frequency of any panel data set is unlikely to coincide with the timing of actual decision making. Manski (1993) has made similar points and states that "a researcher must maintain the hypothesis that the transmission of social patterns really follows the assumed temporal pattern. But empirical studies typically provide no evidence for any particular timing (p540)."

Thus, for estimation I make an additional assumption about the sampling process. That is, I assume that the smoking profile at each school is distributed according to the steady-state distribution P^*, which describes the interaction process. The steady-state distribution can be compared with the empirical cross-sectional distribution of smoking choices across schools.[14] It is described formally as follows. Given data on the smoking choices y_i and background characteristics $\mathbf{x}_i$ of all N_s persons, I make the following assumption about the data generation process. Let $P_s \equiv \{\mathrm{Prob}(\mathbf{y}_s = \boldsymbol{\omega}_s | \mathbf{x}_s) : \boldsymbol{\omega}_s \in \Omega_s\}$ 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.

The steady-state distribution of the interaction process, P^*, can be used as a likelihood function to estimate the model by using cross-section data. Given the assumptions above, the likelihood of a smoking profile, $\mathbf{y}_s$, can be derived as follows. 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,

12. 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.

13. Of course, one can actually model the dynamic smoking behavior differently with panel data. Such a model might be estimable without knowledge of the "order of moves" of persons.

14. A similar approach to estimation has been used by Topa (2001).

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 P^* is given by equation (2.6) in the theorem, the likelihood of $\mathbf{y}_s$ conditional on $\mathbf{x}_s$ at school s is given by

$$P^*(\mathbf{y}_s = \boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) = \exp Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) \Big/ \sum_{\boldsymbol{\eta}_s \in \Omega_s} \exp Q(\boldsymbol{\eta}_s | \mathbf{x}_s, \boldsymbol{\theta}). \tag{3.9}$$

Similarly to equation (2.7), one obtains

$$Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta}) = \frac{1}{2} \sum_i \omega_i (\alpha + \mathbf{x}_i \boldsymbol{\beta}) + \frac{1}{4} \sum_g \sum_{g'} \cdot \sum_i \sum_j \rho_{gg'} \omega_{ig} \omega_{jg'} / N_s, \tag{3.10}$$

where ω_{ig} indicates the choice made by person i of type g. Then, $\sum \omega_{ig}\omega_{jg'}$ is obtained for any pair of individuals, i and j, who belong to types g and g', respectively. Therefore, one can estimate the structural parameter $\boldsymbol{\theta}$ by maximum likelihood. The contribution to the log likelihood by school s is given by

$$\ell_s(\boldsymbol{\theta}) = Q(\mathbf{y}_s | \mathbf{x}_s, \boldsymbol{\theta}) - \log \sum_{\boldsymbol{\eta}_s \in \Omega_s} \exp Q(\boldsymbol{\eta}_s | \mathbf{x}_s, \boldsymbol{\theta}). \tag{3.11}$$

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 $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$.

Concerning the data structure, I assume a cluster sampling scheme in which a large number of peer groups are randomly selected, and in which each peer group is relatively small. Thus the asymptotic analysis is with the number of peer groups S becoming larger, while the size of each peer group N_s is finite. Note that $N = \sum_{s=1}^{S} N_s$.

The estimation strategy described in this paper differs from that used in the discrete-choice model of Brock and Durlauf (2001), who use standard logistic regression to estimate the model parameters under the assumption that each person responds to an identical expectation of others' choices. While this approach may be appropriate in many situations, it is not applicable in the context of the problem analyzed in this paper. The premise of Brock and Durlauf is that the size of the peer group is so large that the individual expectations of others' choices can be approximated by the population average of the observed choices. However, peer groups are not necessarily large. This paper assumes that the peer group comprises students who take the classes of required subjects within the same school cohort. Thus, the large-limit approximation used by Brock and Durlauf might not be applicable to the moderately sized peer groups studied in this paper. Although the validity of the assumption should be judged on a case-by-case basis, the estimation framework used in this paper seems to suit moderately sized peer groups. On the other hand, the framework of Brock and Durlauf could be used when there is one large peer group (e.g., involving interactions in a large neighborhood).

3.3. *Identification of Peer Effects*

In this subsection, I investigate parameter identification.

I begin by showing that *not all* cross-type peer effects can be separately identified from the data. In equation (3.10), note that $\sum_i \sum_j \omega_{ig}\omega_{jg'} = \sum_i \sum_j \omega_{ig'}\omega_{jg}$ for any $g \neq g' \in \{1, \cdots, G\}$, so that

$$\sum_g \sum_{g'} \rho_{gg'} \sum_i \sum_j \omega_{ig}\omega_{jg'} = \sum_{g<g'} (\rho_{gg'} + \rho_{g'g}) \sum_i \sum_j \omega_{ig}\omega_{jg'}. \tag{3.12}$$

This shows that any combination of the parameters, $\rho_{gg'}$ and $\rho_{g'g}$, given the restriction $\rho_{gg'}+\rho_{g'g}$, implies the same likelihood for equation (3.9). Therefore, the parameters $\rho_{gg'}$ and $\rho_{g'g}$ cannot be identified at the same time. Thus, in the empirical analysis, I report $\overline{\rho_{gg'}} = (\rho_{gg'}+\rho_{g'g})/2$. This composite parameter $\overline{\rho_{gg'}}$ can be interpreted as the *average* of the cross-type peer effects between different types of g and g'.

Next, I consider the variations in the data required to identify the peer effects. For simplicity, I consider a model with one type of person ($G = 1$), so that the peer effect, defined only for the within type, is represented by a single parameter, ρ. The generalization to multiple types ($G > 1$) is quite straightforward. Let $\boldsymbol{\mu}(\boldsymbol{\omega})$ be the vector of functions of choices, defined by

$$\boldsymbol{\mu}(\boldsymbol{\omega}) \equiv \left[\frac{1}{S}\sum_{s=1}^{S}\overline{\omega}_s, \quad \frac{1}{S}\sum_{s=1}^{S}\overline{(\omega\mathbf{x})}_s, \quad \frac{1}{S}\sum_{s=1}^{S}\overline{\omega}_s^2 \right],$$

where $\overline{\omega}_s = \sum_i \omega_i/N_s$, $\overline{(\omega\mathbf{x})}_s = \sum_i(\omega_i\mathbf{x}_i)/N_s$ and $\overline{\omega}_s^2 = [\sum_i \omega_i/N_s]^2$ are the average statistics within school s, and all summands are taken over the set of persons I_s at school s.

According to the theory of exponential families (see Lehmann and Casella 1998), $\boldsymbol{\mu}(\boldsymbol{\omega})$ is the canonical sufficient statistic for the parameter $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \rho)$ of the distribution P^*. The maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, is given by the solution to the following system of nonlinear equations:

$$\mathbf{E}\left[\boldsymbol{\mu}(\boldsymbol{\omega})\middle|\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\rho}\right] = \boldsymbol{\mu}(\mathbf{y}), \tag{3.13}$$

where $\mathbf{E}[\boldsymbol{\mu}(\boldsymbol{\omega})]$ is the expected value of $\boldsymbol{\mu}(\boldsymbol{\omega})$ under the P^*, and $\boldsymbol{\mu}(\mathbf{y})$ is the realized value of $\boldsymbol{\mu}(\boldsymbol{\omega})$ observed in the data. Therefore, a set of parameters, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \rho)$, is identified if it is the *unique* solution to equation (3.13).

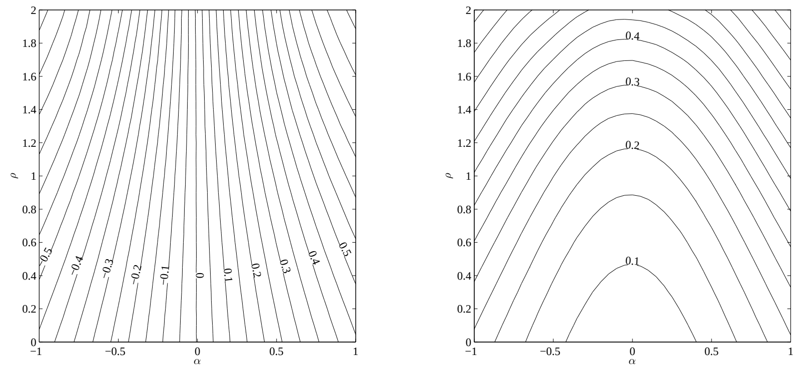
Unfortunately, the expectation in equation (3.13) is a complicated function of the parameter. One cannot solve the system of equations analytically. Although parameter identification cannot be demonstrated analytically in this case, numerical methods can be used to illustrate identification.[15] That is, I provide a map between a set of parameters, $(\alpha, \boldsymbol{\beta}, \rho)$, and the expected canonical statistic, $\mathbf{E}[\boldsymbol{\mu}(\boldsymbol{\omega})]$, which is numerically approximated by using values simulated from the interaction process.

Figure 1 and 2 presents representative selections of contours of the empirical average of $\boldsymbol{\mu}(\boldsymbol{\omega})$ for a grid of a reasonable range of parameters.[16] For simplicity, I consider a model with only one explanatory variable, $\mathbf{x}_s$, which is assumed to be normally distributed across schools $s = 1, \cdots, S$. Figure 1 shows the contours of $\frac{1}{S}\sum_s \overline{\omega}_s$ and $\frac{1}{S}\sum_s \overline{\omega}_s^2$ drawn for $\alpha \in [-1\ 1]$ and $\rho \in [0\ 2]$ given $\beta = 0.5$. Figure 2 shows the contours of $\frac{1}{S}\sum_s \overline{(\omega\mathbf{x})}_s$ and $\frac{1}{S}\sum_s \overline{\omega}_s^2$ drawn for $\beta \in [0\ 1]$ and $\rho \in [0\ 2]$ given $\alpha = 0.5$. [17] In both figures, the intersections between these two sets of contours become singletons over the examined parameter regions. Therefore, these numerical simulation results strongly indicate that the parameters can be uniquely identified from the data.

15. The approach that I adopt in this paper is similar to that used by Conley and Topa (2003), who use simulation exercises to demonstrate the local identification of a dynamic local interaction model.

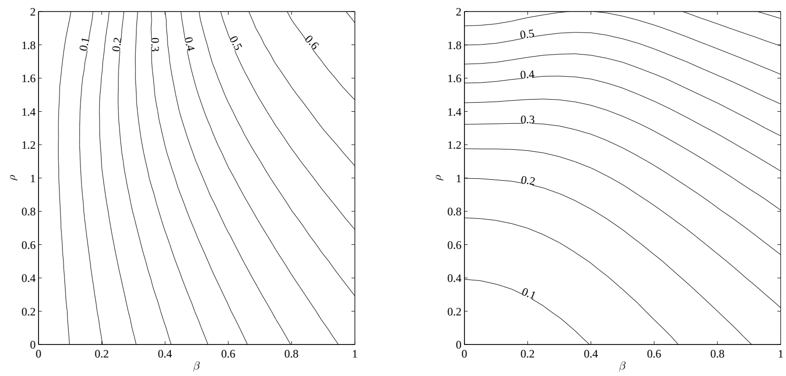
16. The range of parameters is chosen to cover reasonable smoking rates across schools. The smoking choices are simulated for the grid of the values for all parameters with a step size of $\triangle = 0.05$. Detailed implementation strategies of simulation are found in the working paper version, and are also available from the supplementary web site.

17. Different sets of contours drawn for different combinations of parameters can be obtained from the author upon request. The patterns are quite similar to those presented in Figure 1 and 2.



(a) simulated average of $\frac{1}{S}\sum_s \overline{\omega}_s$ (b) simulated average of $\frac{1}{S}\sum_s \overline{\omega}_s^2$

FIGURE 1
Set of contours of the simulated averages of canonical statistics for (α, ρ) in the range of $\alpha \in [-1\ 1]$ and $\rho \in [0\ 2]$ for $\beta = 0.5$



(a) simulated average of $\frac{1}{S}\sum_s (\overline{\omega \mathbf{x}})_s$ (b) simulated average of $\frac{1}{S}\sum_s \overline{\omega}_s^2$

FIGURE 2
Set of contours of the simulated averages of canonical statistics for (β, ρ) in the range of $\beta \in [0\ 1]$ and $\rho \in [0\ 2]$ for $\alpha = 0.5$

The identification exercise by the numerical simulation above is examined only for the limited region of the parameter space, and thus it is local in nature. Yet, it can provide an idea of how and why the peer effect is identifiable from data. To gain the intuition, consider the variance of the average smoking rates between schools, which is given by $\mathrm{Var}(\overline{\omega}_s) = \frac{1}{S}\sum_{s=1}^{S} \overline{\omega}_s^2 - \left[\frac{1}{S}\sum_{s=1}^{S} \overline{\omega}_s\right]^2$. In Figure 1(b), the higher values of $\frac{1}{S}\sum_s \overline{\omega}_s^2$ (located in higher positions) tend to yield the higher value of ρ at the intersection of a given contour of $\frac{1}{S}\sum_s \overline{\omega}_s$, which is the almost vertical line in Figure 1(a). Similarly, in Figure 1(a), the lower absolute values of $\frac{1}{S}\sum_s \overline{\omega}_s$ (located in more central positions) tend to yield the higher values of ρ at the intersection of a given contour of $\frac{1}{S}\sum_s \overline{\omega}_s^2$, which has an inverse-U shape in Figure 1(b). Since the cross-school variance, $\mathrm{Var}(\overline{\omega}_s)$, is related to the difference of the (squared) canonical statistics examined above, this inspection shows that a stronger peer effect, ρ, is indicated *revealed* by a larger $\mathrm{Var}(\overline{\omega}_s)$.

The fundamental idea is that too large a variance of aggregates explained by differences in fundamentals between groups indicates strong peer effects. So, one could use the cross-school variance of average smoking choices to identify the magnitude of peer effects in the estimation. A similar empirical idea identifying peer effects using the cross-group variation is presented in other empirical works on social interactions. [18] In principle, as Glaeser and Scheinkman (2001) states, this empirical approach does not suffer from the reflection problem for identification because it explicitly incorporates the fact that all individuals affect each other. Formally, this approach attributes the sizable variance of aggregates between groups to strong peer effects after controlling for the observable differences in characteristics.

In contrast, in a standard cross-sectional model that excludes peer effects, the interpretation of excessive cross-school variation would be different. The standard model, which does not incorporate an amplifier mechanism provided by peer effects, ascribes the large variation solely to the difference in fundamentals. If the difference in fundamentals is magnified by peer effects, the standard model would overstate the impact of the fundamentals. Thus, in standard models that omit peer effects, the estimate of the coefficient β is larger than that obtained from a social interaction model with peer effects. For example, if the same data were used, the estimated tax elasticity would be larger in the standard model than in the social interaction model of this paper. This is shown in a subsequent section.

3.4. *Monte Carlo Maximum Likelihood*

In practice, the log likelihood function $\ell(\boldsymbol{\theta})$ requires the computation of a normalizing constant term, $\sum_{\boldsymbol{\eta}\in\Omega} \exp Q(\boldsymbol{\eta}|\mathbf{x},\boldsymbol{\theta})$, which is 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.

To circumvent this computational impossibility, 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 $\boldsymbol{\psi} \in \Theta$. Since the Markov chain of the interaction process is irreducible and aperiodic, the steady-state distribution, $P^*(\boldsymbol{\omega}|\mathbf{x},\boldsymbol{\psi})$, is unique. 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 $P^*(\boldsymbol{\omega}|\mathbf{x},\boldsymbol{\psi})$.

18. See, for example, Glaeser et al. (1996) , Glaeser and Scheinkman (2000, 2001), Topa (2001).

Define the following function for an arbitrary fixed parameter $\boldsymbol{\psi}$:

$$\ell_T(\boldsymbol{\theta};\boldsymbol{\psi}) = Q(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) - \log \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\exp Q(\mathbf{y}(t)|\mathbf{x},\boldsymbol{\theta})}{\exp Q(\mathbf{y}(t)|\mathbf{x},\boldsymbol{\psi})}\right], \tag{3.14}$$

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 $\boldsymbol{\psi}$. The likelihood functions are approximated by using simulated samples obtained from the interaction process introduced in section 2. I follow a heuristic approach, proposed by Geyer and Thompson (1992), to determine the value of $\boldsymbol{\psi}$ with which the interaction process is simulated.[19]

4. DATA DESCRIPTION

The main data set used for estimation is the 2000 NYTS. The survey was conducted by the American Legacy Foundation in collaboration with the Centers for Disease Control and Prevention (CDC) Foundation during the spring semester of the academic year 2000.[20] 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.[21] Second, the samples are also restricted to students in schools for which the 2000 *NYTS* collects at least ten male and female students.[22]. 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 with a standard deviation of 24.823 students. Maximum and minimum sample sizes per school are 193 and 23, respectively.

19. Detailed implementation strategies of simulation and estimation are found in the working paper version, and are also available from the supplementary web site.

20. The publicly available data set and codebook can be obtained from the website of the American Legacy Foundation. See also Center for Disease Control and Prevention (2001) for an overview of the results from this survey.

21. This led to $1,153$ students (3.22 percent) being dropped.

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

TABLE 1

Smoking Rates (Percentages) by Gender, Grade and Race Subgroups in Sample

	Male	Female	Middle School	High School	Total
White	21.829	22.192	10.809	30.525	22.012
Black	15.499	11.153	9.850	16.984	13.158
Hispanic	18.111	13.507	10.852	20.730	15.745
Asian	17.237	11.425	5.684	19.786	14.486
Total	19.874	18.059	10.415	26.228	18.951

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools.

Most schools contain either middle-school students (6th–8th grade) or high-school students (9th–12th grade). The number of schools that includes only middle-school students is 126 (41.31 percent) whereas the number of schools that includes only high-school students is 149 (48.85 percent). On the other hand, only 30 schools (9.84 percent) contain both middle-school and high-school students at the same time.

It is found that almost one third of schools (99 schools) contain students from only one racial group. Most of these racially uniform schools contain only white students. In addition, the majority-race proportion exceeds 80 percent in more than half of the schools. Thus, most of the sampled schools are racially segregated, rather than racially integrated.

A dichotomous measure of smoking was constructed for all respondents.[23] 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 nonsmoker, otherwise the student was classified as a current smoker. [24] Table 1 reports the percentages of current smokers for various categories. As expected, the smoking prevalence is higher among high- school students than among middle-school students. Table 1 also shows that smokers are not distributed equally between groups: that is, white youths smoke more than Hispanic youths, who in turn smoke more than black and Asian youths. This pattern has been also found in recent national surveys (e.g., the *Monitoring the Future Survey*).

It is also found that smoking rates vary greatly between schools. The mean of the average smoking rate is 18.52 with a standard deviation of 11.80. The highest smoking rate is about 60 percent and the lowest is close to zero.

Table 2 presents information on the frequency and intensity of cigarette smoking. As shown in Table 2, most students sampled did not have a high level 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 on 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. However, Table 2 also shows that some students were hooked on cigarettes. More than 25 percent of smokers are daily smokers. About one third of the daily smokers are

23. For compatibility with the behavior model described above, the binary smoking-choice variable is set to $y_i = +1$ if person i is a current smoker, and otherwise is set to $y_i = -1$.

24. 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

TABLE 2

Smoking Frequency and Intensity

intensity: number of cigarettes per day	frequency: number of smoking days per month 0 days	1-2 days	3-9 days	10-29days	30 days	total	(percentage)
zero	23728	0	0	0	0	23728	(82.83)
less than one	0	764	214	47	5	1030	(3.60)
from one to five	0	628	978	998	507	3111	(10.86)
from six to ten	0	22	59	177	520	778	(2.72)
from eleven to twenty	0	2	14	61	401	478	(1.67)
more than twenty	0	0	3	6	136	145	(0.51)
total	23728	1416	1268	1289	1569	28647	
(percentage)	(82.83)	(4.94)	(4.43)	(4.50)	(5.48)		

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. All respondents were also asked about the frequency and intensity of their cigarette smoking.

considered as “heavy smokers”, who smoke more than half a pack of cigarettes every day. Thus, a small proportion of students might be considered as potential “addicts”, while the majority of students are considered as “chippers”, smokers who are just beginning to get addicted.

After investigating the frequency and intensity of the relationships between *smokers* for middle-school students and high school students, respectively, I found that, as expected, potential addicts appear to be far more concentrated among high-school students than among middle-school students. In middle schools, only 14 percent of smokers are daily smokers and 6 percent of smokers smoke more than half a pack of cigarettes per day. In high schools, on the other hand, 33 percent of smokers are daily smokers and about one third of daily smokers smoke more than half a pack of cigarettes every day. These results suggests that middle-school and high-school students might have different levels of cumulative past cigarette consumption.

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. 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. [25] Descriptive statistics for these control variables are presented in Table 3.

The second set of control variables reflects the characteristics of the neighborhoods in which the students live. Since the 2000 NYTS data do not provide any information on 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 Offence 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. [26]

25. Cigarette taxes at 31 May are used because the 2000 NYTS was carried out during the spring semester. The data set is from Orzechowski and Walker (2001).

26. All the county variables used for estimation are available from the supplementary web site.

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

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. Each student was asked the following question about school-based prevention programs (1) school program1: did you practice ways to say "No" to tobacco in any of your classes? (2) school program 2: were you taught in any of your classes the reasons why people your age smoke? (3) school program 3: were you taught in any of your classes that most people your age do not smoke cigarettes? (4) school program 4: were you taught in any of your classes about the effects of smoking, like it makes your teeth yellow, causes wrinkles, or makes you smell bad?

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. All estimates are calculated with a final Monte Carlo sample size of $10,000$ experiments.[27] The estimated parameters from the benchmark model of youth smoking behavior are reported in Tables 4 and 5. The point estimates of the county variables are available from the supplementary web site.

Table 4 reports the estimated coefficients of the background characteristics. For comparison, estimates from standard logistic regression are presented in the first column of the table. No significant difference in the pattern of coefficients is found between the estimates from standard logistic regression and those from maximum likelihood method, except in relation to the constant term. As shown, almost all variables are significant and have the expected signs. The estimates provide some support for the idea that grade,

27. 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 by using simulations of $T = 30,000$ and $T = 50,000$. The associated estimation results were virtually identical to those obtained from simulations of $T = 10,000$.

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 and Zinman 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 (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.[28]

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.[29] 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.

The estimates of peer effects (ρ_{MM},ρ_{FF},ρ_{MF}) are reported in Table 5. First, standard logistic regression yields greater estimates of peer effects than does the maximum likelihood method. The difference between the two estimators is significant especially for the within-gender peer effects ρ_{MM} and ρ_{FF}. [30] Second, 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 with average smoking prevalence to a new school with one standard deviation from the average. My results predict that in such a case, the probability of smoking would increase by about 12 percentage points (from 14.6 percent to 26.5 percent) when evaluated at the sample mean of the other variables. I found that the elasticity in response to the change of the smoking environment is 1.616. Such 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

28. The estimates of other specifications are available on request.

29. See *Report of the Surgeon General* (1998), chapter 6.

30. The estimation results are consistent with several simulation studies. For example, Geyer (1991), Geyer and Thompson (1992), Huffer and Wu (1998) examine the finite sample properties of different types of estimators, including the standard logistic regression estimator (maximum pseudo likelihood estimator), and the Monte Carlo ML estimator that I employed in this paper. A general conclusion from these simulation studies is that the standard logistic regression estimator tends to overestimate the parameter of endogenous dependency especially when the magnitude of the dependency parameter is large.

TABLE 4

The estimated coefficients of individual background characteristics in the baseline model: smoking choices among students in grades from six to twelve

Parameters Coefficient on	Standard Logit (1)	MonteCarlo MLE (2)	(3)	(4)
Constant	-1.04969	-0.68177	-0.03845	-0.87633
	(4.60251)	(3.69140)	(3.53725)	(3.58852)
7th Grade	0.47629	0.47755	0.46843	0.46602
	(0.09070)	(0.08802)	(0.08763)	(0.08605)
8th Grade	0.95070	0.94492	0.93562	0.93465
	(0.08491)	(0.08213)	(0.08088)	(0.08058)
9th Grade	0.92397	0.98692	0.97299	0.99142
	(0.08942)	(0.08437)	(0.08285)	(0.08340)
10th Grade	1.04231	1.07936	1.06156	1.07325
	(0.09147)	(0.08838)	(0.08718)	(0.08789)
11th Grade	1.05953	1.07958	1.05034	1.06054
	(0.09211)	(0.09053)	(0.08826)	(0.08822)
12th Grade	1.12002	1.15153	1.13572	1.15289
	(0.096140	(0.09397)	(0.09157)	(0.09232)
Asian	-0.55555	-0.54672	-0.54873	-0.54734
	(0.13346)	(0.13176)	(0.12929)	(0.12993)
Black	-0.83468	-0.80229	-0.80005	-0.79480
	(0.07546)	(0.07247)	(0.07182)	(0.07200)
Hispanic	-0.35979	-0.39855	-0.41438	-0.42327
	(0.07281)	(0.07055)	(0.06743)	(0.06971)
White Male	-0.10788	-0.12576	-0.14919	-0.13157
	(0.07527)	(0.06984)	(0.06733)	(0.06944)
Asian Male	0.51544	0.47027	0.43158	0.44862
	(0.17996)	(0.17744)	(0.17336)	(0.17580)
Black Male	0.27928	0.24751	0.20722	0.27635
	(0.11708)	(0.11557)	(0.11408)	(0.11248)
Hispanic Male	0.24973	0.22300	0.18893	0.25128
	(0.11311)	(0.10729)	(0.10414)	(0.10617)
Weekly Income	0.02892	0.02878	0.02903	0.02911
	(0.00232)	(0.00231)	(0.00224)	(0.00226)
Working Dummy	0.31307	0.31798	0.32408	0.32637
	(0.04269)	(0.04204)	(0.04144)	(0.04216)
Smokers in Home	0.95056	0.96354	0.96556	0.96087
	(0.03326)	(0.03308)	(0.03181)	(0.03222)
See Actors Smoking in TV	0.05993	0.05762	-	0.07542
	(0.04997)	(0.04963)	-	(0.04378)
See Actors Smoking in Movie	0.06421	0.05875	0.06059	-
	(0.05683)	(0.05656)	(0.04968)	-
School Program 1	-0.19578	-0.19903	-0.20304	-0.18129
	(0.04612)	(0.04626)	(0.04410)	(0.04499)
School Program 2	-0.08713	-0.09634	-0.10557	-0.11811
	(0.04440)	(0.04450)	(0.03855)	(0.03880)
School Program 3	0.14560	0.14327	0.16336	0.14783
	(0.04715)	(0.04733)	(0.04570)	(0.04590)
School Program 4	-0.00596	-0.01018	-	-
	(0.04176)	(0.04179)	-	-
Cigarette Tax	-0.21504	-0.23608	-0.21771	-0.20983
	(0.11475)	(0.09212)	(0.08868)	(0.08925)

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. Asymptotic standard errors are in parenthesis.

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.

Table 5 shows that the magnitudes of ρ_{MM} and ρ_{FF} are not significantly different,

TABLE 5

The estimated peer effects in the baseline model: smoking choices among students in grades from six to twelve

Peer Effects	Standard Logit	MonteCarlo MLE		
	(1)	(2)	(3)	(4)
ρ_{MM}	2.19279*	1.70269*	1.66938*	1.66987*
	(0.21248)	(0.20653)	(0.19879)	(0.20741)
ρ_{FF}	2.00905*	1.65828*	1.73793*	1.628*
	(0.20660)	(0.20407)	(0.19796)	(0.20646)
ρ_{MF}	0.76445*	0.7275*	0.74057*	0.76848*
	(0.16775)	(0.16668)	(0.16465)	(0.17157)

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

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.

If there is race homogeneity in peer groups, one would expect peer interaction to differ between racial groups. To explore the possibility, I estimated peer effects that are assumed to be constant within racial groups but different between racial groups. The race-based peer effects are denoted by $(\rho_{WW}, \rho_{BB}, \rho_{HH}, \rho_{WB}, \rho_{BH}, \rho_{HW})$, where the subscripts represents each race group $\{W, B, H\}$. One drawback with the estimation of race-based peer effects is that this requires further subsampling of the data set. As stated earlier, it is found that many schools are racially segregated. The 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.

The estimates of the race-based peer effects are given in Table 6. I start with estimates from the logit model for the purposes of comparison.[31] Once again the estimates of peer effects from standard logistic regression tend to be greater than those from the maximum likelihood method. The difference is significant for within-male and within-black peer effects.

Table 6 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

31. 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 6

Estimated raced-based peer effects: smoking choices among students in grades from six to twelve

Peer Effects	Standard Logit (1)	MonteCarlo MLE (2)	(3)	(4)
ρ_{WW}	1.69914*	1.55833*	1.5875*	1.53104*
	(0.21748)	(0.20126)	(0.19402)	(0.20347)
ρ_{BB}	1.21111*	0.77466*	0.79752*	0.72025*
	(0.36105)	(0.36861)	(0.36423)	(0.36913)
ρ_{HH}	1.48881*	1.4452*	1.39753*	1.45932*
	(0.37557)	(0.36970)	(0.35295)	(0.35689)
ρ_{WB}	0.09463	0.01568	0.03009	-0.00037
	(0.25670)	(0.25102)	(0.25006)	(0.25667)
ρ_{WH}	0.63648*	0.71963*	0.75615*	0.71113*
	(0.28831)	(0.27574)	(0.26759)	(0.28123)
ρ_{BH}	0.91066*	0.99595*	0.94710*	0.88158*
	(0.33225)	(0.33315)	(0.33104)	(0.34128)

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students, and there are more than two racial groups. The sample size is $N = 13622$ from $S = 156$ schools. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

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 uniformly positive. This suggests that there are no, or negligible, peer interactions between white and black students in terms of smoking decisions. If white and black students hardly interact with each another, as the results suggest, there is 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 necessarily 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 6 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}. As shown, this tendency is clear among white and Hispanic students. This is illustrated in Table 6 by values for the within-race peer effects, ρ_{WW} and ρ_{HH}, that are substantially larger than those 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 County 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 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 added to the model fixed effects that explicitly account for unobserved heterogeneity between counties. Under the assumption that all unobserved heterogeneities vary at the county level, the endogenous peer effect is separately identified from the omitted factors. The county fixed effects account for a common unobserved factor confronted by everyone living in the same county.

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.[32] 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 $\delta = (\delta_1, \cdots, \delta_L)$. I assume that the systematic utility of person i in county ℓ is given by

$$b_i(\mathbf{x}_i) = \delta_\ell + \mathbf{x}_i \boldsymbol{\beta}. \tag{5.15}$$

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 nonsmoking, $y_i = -1$, is chosen if $y_i^* < 0$. The space-specific factor δ_ℓ yields neighborhood correlation if it is not taken into account. In that case, the unobserved error term, $(\delta_\ell + \epsilon_i)$, is correlated between individuals who belong to the same county due to the common factor δ_ℓ.

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 $\mathbf{y}_\ell \in \Omega_\ell$. Suppose that each person $i \in I_\ell$ belongs to one of the S_ℓ schools in the county. Given that $I_\ell = (I_1, \cdots, I_s, \cdots, I_{S_\ell})$, $\Omega_\ell = \prod_{s=1}^{S_\ell} \Omega_s$. For the parameter $\boldsymbol{\theta} = (\delta, \boldsymbol{\beta}, \boldsymbol{\rho})$, the log likelihood function of county ℓ is given by

$$\ell_\ell(\boldsymbol{\theta}) = Q(\boldsymbol{\omega}_\ell | \mathbf{x}_\ell, \boldsymbol{\theta}) - \log \sum_{\boldsymbol{\eta}_\ell \in \Omega_\ell} \exp Q(\boldsymbol{\eta}_\ell | \mathbf{x}_\ell, \boldsymbol{\theta}), \tag{5.16}$$

where $Q(\boldsymbol{\omega}_\ell | \mathbf{x}_\ell, \boldsymbol{\theta}) = \sum_{s=1}^{S_\ell} Q(\boldsymbol{\omega}_s | \mathbf{x}_s, \boldsymbol{\theta})$ analogously to equations (3.10) and (3.11). The overall log likelihood is $\ell(\boldsymbol{\theta}) = \sum_\ell \ell_\ell(\boldsymbol{\theta})$, and the maximum likelihood estimator is $\hat{\boldsymbol{\theta}} \equiv \arg\max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$.

The following assumption is made concerning the sampling structure for the fixed-effects model. According to the cluster sampling scheme that I assumed above, the number of schools S is large in the empirical analysis. Given the two-stage sampling design adopted for the NYTS, I assume that more and more observations are taken for S by increasing the number of counties L, while each country contains several school units S_ℓ. Note that $S = \sum_{\ell=1}^{L} S_\ell$. [33]

One potential problem with the maximum likelihood method described above is that it is inconsistent when $S \to \infty$. The sampling scheme for the fixed-effects model leads to an increasing number of counties L as the number of schools S become large.

32. There are $L = 146$ counties in the sample.

33. The NYTS employs the following multistage cluster sample design. The first-stage sampling contained about 150 primary sampling units, each comprising a large county or a group of smaller adjacent counties. At the second sampling stage, several schools were selected from each primary sampling unit with a probability proportional to weighted school enrollment.

It thus follows that the number of fixed effects, which attempts to capture unobserved county-specific factors, grows to infinity. This is the well-known "incidental parameter" problem.

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 sufficient statistic of δ_ℓ is $\tau_\ell = \sum_{i \in I_\ell} y_i$, which is the total number of smokers in county ℓ.[34] Define all possible states of the smoking profile given the restriction that the number of smokers is fixed at τ_ℓ as follows:

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

As shown in the appendix, the conditional probability, P^* given τ_ℓ, is

$$P^*(\boldsymbol{\omega}_\ell | \mathbf{x}_\ell, \boldsymbol{\theta}; \tau_\ell) = \exp Q_1(\boldsymbol{\omega}_\ell | \mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) \Big/ \sum_{\boldsymbol{\eta}_\ell \in B_\ell} \exp Q_1(\boldsymbol{\eta}_\ell | \mathbf{x}_{1\ell}, \boldsymbol{\theta}_1). \tag{5.17}$$

In this case, it can be shown that the Q_1 function does not depend on the nuisance parameter, δ_ℓ, as follows:

$$Q_1(\boldsymbol{\omega}_\ell | \mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) = \sum_{s=1}^{S_\ell} \left[\frac{1}{2} \sum_i \omega_i \mathbf{x}_{1i} \boldsymbol{\beta}_1 + \frac{1}{4} \sum_g \sum_{g'} \cdot \sum_i \sum_j \rho_{gg'} \omega_{ig} \omega_{jg'} / N \right]. \tag{5.18}$$

Thus, the conditional likelihood is given by

$$\ell_{1\ell}(\boldsymbol{\theta}_1) = Q_1(\boldsymbol{\omega}_\ell | \mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) - \log \sum_{\boldsymbol{\eta}_\ell \in B_\ell} \exp Q_1(\boldsymbol{\eta}_\ell | \mathbf{x}_{1\ell}, \boldsymbol{\theta}_1) \tag{5.19}$$

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 $\mathbf{x}_{1\ell}$. Note that the summation in the normalizing constant covers the states in the smoking profile that satisfy the restriction given by τ_ℓ.

Because of computational difficulties,[35] the conditional log likelihood function, $\ell_{1\ell}(\boldsymbol{\theta}_1)$, is obtained by using the Monte Carlo technique described in section 3. By analogy to equation (3.14), I define a Monte Carlo conditional likelihood function for an arbitrary

34. As the overall likelihood is given by $\ell(\boldsymbol{\theta}) = \sum_\ell \ell_\ell(\boldsymbol{\theta})$, the Q-function of the likelihood $\ell(\boldsymbol{\theta})$ contains the following term (see the argument in subsection 3.3 on parameter identification) $\sum_\ell \delta_\ell \left[\sum_{s=1}^{S_\ell} \bar{\omega}_s \right]$. Here the $\bar{\omega}_s$ is the sum of the choices of all persons in school s. One can thus interpret the term $\sum_{s=1}^{S_\ell} \bar{\omega}_s$ as the summation of choices for all persons in county ℓ. Using the definition of the τ_ℓ, one finds $\sum_{s=1}^{S_\ell} \bar{\omega}_s = \sum_{i \in I_\ell} \omega_i \equiv \tau_\ell$. Therefore, it is shown that the Q-function of the likelihood $\ell(\boldsymbol{\theta})$ contains $\sum_\ell \delta_\ell \tau_\ell$. Following the theory of exponential families, one can find that the τ_ℓ is the sufficient statistic for δ_ℓ.

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

fixed parameter $\boldsymbol{\psi}_1$ as follows:

$$\ell_{1T}(\boldsymbol{\theta}_1; \boldsymbol{\psi}_1) = Q_1(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1) - \log \left[\sum_{t=1}^{T} \frac{\exp Q_1(\mathbf{y}(t)|\mathbf{x}, \boldsymbol{\theta}_1)}{\exp Q_1(\mathbf{y}(t)|\mathbf{x}, \boldsymbol{\psi}_1)} \right], \tag{5.20}$$

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 (5.17). It should be noted, however, that the interaction process used to compute the Monte Carlo *unconditional* log likelihood function (i.e., equation (3.14)) 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 $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1)$. In Appendix, I present an example of the Markov chain that has the conditional distribution $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1; \tau_\ell)$ as its steady-state distribution. The Markov chain, which is aperiodic and irreducible, is convergent to the $P^*(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell, \boldsymbol{\theta}_1; \tau_\ell)$ as $t \to \infty$.

I estimate the structural parameter, $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1, \boldsymbol{\rho})$, by 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.

Table 7 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 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 7.

Table 8 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 8 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 8 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, by using the model with county fixed effects. The results are reported in Table 9. These point estimates are directly comparable with those presented in Table 6. The standard errors of these estimates are larger, which indicates that the point estimates are not as precise as those in Table 6. This may explain why not all raced-based peer effects are statistically significant in Table 9. The estimated peer effects based on the inclusion of the fixed effects are smaller than those in Table 6. 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 8 and 9) are qualitatively similar to those from the model without the fixed effects (Table 5 and 6). The estimation results show that the inclusion of county-specific fixed

TABLE 7

The estimated coefficients of individual background characteristics in the fixed-effects model: smoking choices among students in grades from six to twelve

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

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. All parameters are estimated by Monte Carlo maximum likelihood method. Asymptotic standard errors are in parenthesis.

effects does not negate 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. Thus, as long as one is willing to accept that most of the omitted variables vary only at the county level, the evidence can be interpreted in favor of positive

TABLE 8

The estimated coefficients of peer effects in the fixed-effects model: smoking choices among students in grades from six to twelve

Peer Effects	MonteCarlo MLE (2)	(3)	(4)
ρ_{MM}	1.5209*	1.47377*	1.54619*
	(0.24665)	(0.25002)	(0.25164)
ρ_{FF}	1.4803*	1.60347*	1.46383*
	(0.24326)	(0.23562)	(0.24089)
ρ_{MF}	0.52547*	0.56642*	0.5856*
	(0.20739)	(0.20826)	(0.2062)

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. All parameters are estimated by Monte Carlo maximum likelihood method. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

TABLE 9

The estimated coefficients of race-based peer effects in the the fixed-effects model: smoking choices among students in grades from six to twelve

Peer Effects	MonteCarlo MLE (2)	(3)	(4)
ρ_{WW}	1.49926*	1.50204*	1.47278*
	(0.31725)	(0.30838)	(0.31973)
ρ_{BB}	0.7817*	0.73457*	0.63131*
	(0.53012)	(0.52385)	(0.52662)
ρ_{HH}	0.51174	0.51186	0.58596
	(0.55511)	(0.54107)	(0.5332)
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)

Note- Data are composed of Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students, and there are more than two racial groups. The sample size is $N = 13622$ from $S = 156$ schools. All parameters are estimated by Monte Carlo maximum likelihood method. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

peer effects.

It is important to keep in mind that the aforementioned results do not exclude the possibility that some unobserved attributes may be influential. If unobserved "neighborhood" factors vary between areas, smaller than the county level, such as "streets" for example, these estimates of peer effects may still overestimate the total peer effect. I ignore the potential omitted variable bias because there is no sufficient variation in the data that allows for fixed effects to capture systematic differences between subcounty areas. Hence, the estimates presented above might be interpreted as an upper bound of the true peer effects. Nevertheless, the fixed-effects model and the conditional likelihood argument could be applicable without any modification to the case where the unobserved factors are aggregated at the subcounty level, only if data were available. Such an analysis is an interesting direction for future research.

6. DISCUSSION

In the previous section, I presented estimates of peer effects. The results can be interpreted as evidence of peer effects among secondary school students in smoking decisions. This section examines a range of alternative specifications to check the robustness of the preceding results. I then use the parameter estimates of the structural model to simulate the effects of proposed smoking policies.

6.1. *Alternative Specification of Peer Groups*

In the previous section, I assumed that a person's peer group is defined by two characteristics, gender and race. Although many sociological studies suggest that these characteristics are possibly related to peer group composition, the estimates could be sensitive to this specification of peer groups.

One problem with the previous definition of peer groups is that gender and race enter the model in a *linear* form. That is, peer effects are allowed to change on the basis of any one of these two characteristics. In order to mitigate this problem, I estimate the model under a less restrictive assumption. In particular, I allow peer effects to vary by race and gender simultaneously. [36]

The results are presented in Table 10. Peer effects are categorized into four groups, based on whether they are within-group or between-group effects. For the same reason given in the previous section, the sample is restricted to individuals who belong to schools in which there are more than two racial groups. The sample size is $N = 13622$. For all estimates, the standard errors are substantially larger than those of the previous estimates, which is to be expected given the increase in the number of parameters and the decrease in the sample size.

I begin by discussing the estimates from the model without fixed effects, which are reported in the first column of Table 10. This shows that several peer effects are positive and significant. I then add fixed effects. The results are reported in the second column of Table 10. It is interesting that three estimated peer effects that were significant in the model without fixed effects are insignificant when fixed effects are included. Specifically, inclusion of the fixed effects reduces the between-gender peer effects between black and Hispanic students substantially. Since these peer effects are probably weak in practice, the positive estimates obtained when fixed effects are excluded might be spurious because of unobserved common factors.

Table 10 establishes the following three findings. First, within-race peer effects tend to be larger than between-race peer effects. When fixed effects are included, the largest peer effect is found among black male students, and the second largest is found among white female students. On the other hand, no positive and significant peer effects are found for between-race peer effects. In particular, no significant evidence of peer effects between white and black students is obtained. This is consistent with previous findings that there is little interaction between these two racial groups. Second, the evidence on whether within-gender peer effects are stronger than between-gender peer effects is mixed. Comparison of within-race and between-race peer effects is less clear because of the large standard errors of the estimates. Nevertheless, it can be seen that, for black

36. For example, I consider peer effects between male white students and female black students, which may differ, for example, from peer effects between female white students and black male students. However, this modification increases the number of estimated peer effects, from 6 to 21.

TABLE 10

The estimated coefficients of gender and race based peer effects in the model: smoking choices among students in grades from six to twelve

Peer Effects	without fixed effects	with fixed effects
(1) within race; within gender		
white male-white male	1.09718	0.84071
	(0.64132)	(0.72533)
white female-white female	1.99610*	1.97717*
	(0.48082)	(0.54431)
black male-black male	2.05083*	2.08932*
	(1.10483)	(1.21822)
black female-black female	0.45058	0.44938
	(0.98469)	(1.06182)
hispanic male-hispanic male	1.54189*	1.33725
	(0.80785)	(1.08822)
hispanic female-hispanic female	0.36991	-0.60811
	(0.94691)	(1.08198)
(2) within race; between genders		
white male-white female	1.46888*	1.39364*
	(0.46258)	(0.52843)
black male-black female	0.14077	−0.10236
	(0.78197)	(0.86825)
hispanic male-hispanic female	1.93743*	0.72877
	(0.64820)	(0.82115)
(3) betwwen races; within gender		
white male-black male	0.82193	0.29827
	(0.80509)	(0.90117)
white male-hispanic male	1.08192	1.10626
	(0.72201)	(0.90584)
white female-black female	0.36464	0.06603
	(0.62748)	(0.71815)
white female-hispanic female	0.46598	0.02245
	(0.7172)	(0.81785)
black male-hispanic male	0.10215	−0.75752
	(1.26952)	(1.33411)
black female-hispanic female	−1.75191*	−2.88424*
	(0.87552)	(1.00213)
(4) between races; between genders		
white male-black female	−1.06325	−1.25204
	(0.72978)	(0.80465)
white male-hispanic female	0.21252	−0.50348
	(0.78221)	(0.94037)
white female-black male	−0.02553	−0.37501
	(0.73214)	(0.78994)
white female-hispanic male	0.79950	0.22420
	(0.74853)	(0.84582)
black male-hispanic female	3.06901*	1.14289
	(1.09675)	(1.22373)
black female-hispanic male	2.11233*	0.28526
	(1.05631)	(1.12711)

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students, and there are more than two racial groups. The sample size is $N = 13622$ from $S = 156$ schools. All parameters are estimated by Monte Carlo maximum likelihood method. Estimates are based on the best fitting specification (4) in Table 4. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

students, the within-male peer effect is significantly stronger than the between-male-and-female peer effect.[37] Third, a negative and significant peer effect is found between racial groups. Based on the estimates from the fixed-effects model, there is a negative peer

37. For white students, the estimate of the peer effect between male and female students is statistically significant. However, I cannot reject the hypothesis that the peer effect between male and female students is larger than the effect among male students at the 10 percent significance level because these point estimates have large standard errors.

effect between black female and Hispanic female students. This might be evidence of "snob effects" between these groups, in the sense that one group's demand for smoking falls when a person from the other group is smoking. Of course, this evidence is not sufficient to confirm the possibility of nonconforming behavior among these two groups of adolescents, and hence, the issue of negative peer effects warrants further research.

Hence, there is ample evidence of positive and significant peer effects even under alternative specifications of peer groups.[38] Moreover, the finding that within-type peer effects are stronger than between-type peer effects appears to be consistent for all groups. Hence, the previous results are not merely artifacts of specifically defined peer groups.

6.2. *Heterogeneity due to Addictive Stocks*

I have thus far ignored the addictive nature of cigarette smoking by assuming that the accumulation of addictive substances might have negligible effects on youths' smoking decisions. Yet, as Tables 2 show, there is evidence that some students, if only a small fraction, may already be addicted to cigarettes to some extent.

As is well known, addiction is an irreversible process with what is often referred to as the "withdrawal effect". When a stock of an addictive substance (i.e., nicotine) is accumulated beyond a critical threshold, addicts fixated on cigarettes are less likely to quit smoking even within a nonsmoking peer environment. Thus persons with a greater stock of an addictive substance might be less responsive to peer influence than those with a lower stock. Thus, persons with a greater stock of an addictive substance might be *less* responsive to peer influences than those with a lower stock. To explore the possibility of differences in smoking behavior due to different levels of addictive stocks, I estimate smoking models for middle-school and high-school students separately. I assume that high-school students have a greater accumulation of past cigarette consumption than do middle-school students. This assumption seems plausible given the results that heavy smokers are more concentrated in high schools than in middle schools.[39]

To estimate the effects, I split the data set used for the baseline estimates ($N = 29,385$) into two subsamples, middle-school students (6th–8th graders) and high-school students (9th–12th graders). I exclude students in "mixed" schools, which include both middle-school and high-school students, so that neither subsample includes 6th–12th-grade students at the same time.[40] Therefore, in the context of the samples used for estimation, one can assume that middle-school students interact exclusively with middle-school students within the same school, as do high-school students. The effect of addiction will not be transmitted from high-school students to middle-school students because there is no externality between middle- and high-school students. For the final analysis, I used

38. Another concern about the earlier definition of peer groups is that peer groups might be based on aspects other than gender and race. To address the issue, I use student grades to extend the definition of peer groups. Assuming that peer groups are characterized by gender and grade simultaneously, there are six peer-effects parameters to be estimated. The result of estimation is available from the supplementary web site. The result shows that, either within or between groups, the peer effects are significantly positive. Again, it is found that peer effects are stronger within group than between groups. Thus, the conclusion is not changed under the alternative specification of peer groups. For detailed discussion, see the working paper version of the paper.

39. Arguably, to take into full account the effects of addiction on youth smoking decision making, one would need to control for past smoking consumption, or equivalently, the stock of accumulated nicotine in the past. However, the NYTS data set provides only an imprecise measure of past smoking consumption. The quality of the information is questioned because it is collected retrospectively. The teenagers' ability to recollect their past smoking consumption might be limited.

40. This leads to the exclusion of $3,139$ students in 31 "mixed" schools. This is only 10.87 percent of the baseline data set.

TABLE 11

The estimated coefficients of peer effects for middle school students in grades from six to eight and high school students in grades from nine to twelve

Peer Effects	middle school $N = 11829$ $S = 126$	high school $N = 14363$ $S = 149$
ρ_{MM}	2.64823*	1.43518*
	(0.53032)	(0.28923)
ρ_{FF}	1.60864*	1.40478*
	(0.59035)	(0.2798)
ρ_{MF}	0.59959	0.73054
	(0.44537)	(0.23078)

Note- Data are composed of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. Middle school students are students who were in grade 6, 7, or 8, and high school students are students who were in grade 9, 10, 11, or 12 at the time of the survey. All parameters are estimated by Monte Carlo maximum likelihood method. Estimates are based on the best fitting specification (4) in Table 4. Asymptotic standard errors are in parenthesis. Asterisks indicate significance at 5 percent level.

$N = 11,829$ students from $S = 126$ middle schools and $N = 14,363$ students from $S = 149$ high schools for estimation.

Table 11 compares the estimates of gender-based peer effects for the subsamples of middle-school and high-school students.[41] Table 11 reports the estimated parameters from the model without county fixed effects. Because many counties in the data sets tend to contain either one middle school or one high school [42], there is insufficient variation to allow for county-specific fixed effects. Thus, maximum likelihood estimates from the fixed-effects-model cannot be obtained for these subsamples.

The first column reports the estimated peer effects for middle-school students. There is a striking difference between the estimated within-gender peer effects, ρ_{MM} and ρ_{FF}. That is, the peer effect among male middle-school students is much larger than that among female middle-school students. This suggests that male students are more strongly influenced by each other than are female students in the lower (6th–8th) grades. Again, the within-gender peer effects, ρ_{MM} and ρ_{FF}, are larger than the between-gender peer effect, ρ_{MF}. Note that the between-gender peer effect is positive but insignificant. Hence, peer influences are negligible between genders, and middle-school students quite segregated on the basis of gender.

The second column reports the estimated peer effects for high-school students. For higher-grade (9th–12th) students, there is no significant difference between the estimated within-male peer effect, ρ_{MM}, and the within-female peer effect, ρ_{FF}. Moreover, although the estimated between-gender peer effect, ρ_{MF}, is smaller than the within-gender peer effect, it remains positive and significant at the 5 percent level. This result suggests that, although high-school students are more likely to be influenced by friends of the same

41. When I tried to estimate race-based peer effects for the subsamples of middle-school and high-school students, the maximum likelihood estimate did not converge. As stated in the previous section, the estimation of race-based peer effects requires further exclusion of racially uniform schools from the data set. This only leaves less than one fifth of the whole sample for estimation. Given that, as already stated, peer effects are identified by variations in choices between schools, the small sample size is makes reliable estimation problematic. In fact, in the initial search for an appropriate starting value, $\boldsymbol{\psi}$ tended to oscillate widely, and hence, the maximum likelihood estimate, $\boldsymbol{\theta}$, failed to converge.

42. To be more precise, 92 of 143 counties (64.33 percent) have only one middle school and 80 of 125 counties (64.00 percent) have only one high school.

gender than by friends of a different gender, they might take into account cross-gender peer influences. Taken together, these results suggest that, at higher grades, within-gender peer effects diminish whereas between-gender peer effects strengthen. This pattern can be explained by findings from sociology that the degree of gender homogeneity in friendship declines from middle school to high school (e.g., Shrum et al. 1988). Relative to the baseline estimates in Table 5, the magnitudes of the estimates are smaller for all types of peer effects. However, the reduction is not substantial, and all peer effects are remain statistically significant at the 5 percent level.

Are middle-school or high-school students more sensitive to peer pressures to smoke? To answer the question, consider a student of average background characteristics in a school of average smoking prevalence. Suppose that this student relocates from a school with average smoking prevalence to a new school in which the proportion of smokers exceeds the average by one standard deviation. The estimated model for middle-school students predicts that the probability of such a student being smoker increases from 0.085 to 0.180. The elasticity in response to a change in the smoking environment is 2.208. On the other hand, the estimated model for high-school students predicts an increase in the probability of being smoker from 0.209 to 0.338. The elasticity of smoking participation is 1.237. Hence, middle-school students are more responsive to a peer-induced change in the smoking environment than are high-school students.

Overall, these results suggest the following conclusions. First, middle-school and high-school students respond differently to peer pressures. Middle-school students are more influenced by the smoking behavior of their friends than are high-school students. This is compatible with the a priori expectation that addiction might reduce peer effects in youth smoking decisions. Second, and nevertheless, ample evidence of substantial positive peer interactions among both middle-school and high-school students remains. Positive peer effects remain after controlling for differences in the stock of addiction. Hence, there are substantial peer effects for middle-school students and high-school students.

6.3. *Policy Experiments*

To reduce youth smoking, it is essential to know whether proposed smoking policies are effective. In this section, I evaluate the effects of various smoking policies on youth smoking behavior based on the estimated model above. I predict the effects of hypothetical policy changes on smoking rates, and compare these effects with those of past policy interventions of a similar nature.

The first policy experiment relates to cigarette taxes. The first panel of Table 12 reports predicted smoking rates under a variety of hypothetical taxes. To obtain the predicted smoking rates, I repeat experiments by simulating smoking profiles from the social interaction model based on the parameter estimates for the best-fitting specifications.[43] The first column reports simulated tax effects on the smoking rates of students in all grades from 6 through 12. As expected, smoking rates decrease as cigarette taxes increase. The estimated elasticities are presented in square brackets.[44] These elasticities range from −0.18 to −0.20 for the tax increases considered. This implies

43. I use the estimates reported in column (4) of Table 4 and 5 relating to students of all grades, and use the estimates reported in Table 11 relating to middle- and high-school students separately.

44. Each tax elasticity is $(\Delta y/y)/(\Delta\tau/\tau)$, 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.

TABLE 12

Simulated Smoking Prevalence Following Tax Increases

	students of all grades N=29385	middle school students N=11829	high school students N=14363
baseline	18.941	10.304	26.356
A. hypothetical tax impacts on smoking rates; with peer effects			
20 cents	18.084	9.716	25.646
	[-0.202]	[-0.255]	[-0.120]
40 cents	17.264	9.165	24.955
	[-0.198]	[-0.247]	[-0.119]
60 cents	16.483	8.651	24.278
	[-0.193]	[-0.239]	[-0.117]
80 cents	15.738	8.166	23.615
	[-0.189]	[-0.232]	[-0.116]
100 cents	15.024	7.706	22.974
	[-0.185]	[-0.225]	[-0.115]
120 cents	14.344	7.281	22.348
	[-0.181]	[-0.218]	[-0.113]
B. hypothetical tax impacts on smoking rates; without peer effects			
20 cents	0.179	0.097	0.252
	[-0.248]	[-0.269]	[-0.196]
40 cents	0.169	0.091	0.241
	[-0.243]	[-0.262]	[-0.193]
60 cents	0.159	0.085	0.23
	[-0.238]	[-0.256]	[-0.190]
80 cents	0.15	0.08	0.219
	[-0.233]	[-0.249]	[-0.188]
100 cents	0.141	0.075	0.209
	[-0.228]	[-0.243]	[-0.185]
120 cents	0.133	0.07	0.199
	[-0.224]	[-0.237]	[-0.182]

Note- Data are composed of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. Middle school students are between 6th and 8th grades, high school students are between 9th and 12th grades. All simulated values are computed from the average over 1000 experiments. The elasticities are computed based on the average tax of 89.26 cent in sample. Tax elasticities are in blankets.

that a 10 percent increase in federal and state taxes reduces smoking rates among middle-school and high-school students by roughly 2 percent.

The second and third columns of Table 12 present simulated tax effects on the smoking rates of middle- and high-school students, respectively. Again, the tax increases have negative and substantial effects on youth smoking. The results show that the elasticities range from -0.22 to -0.26 for middle-school students, and from -0.11 to -0.12 for high-school students. Thus, middle-school students are much more responsive to tax changes than are high-school students: the tax elasticity for middle-school students is about twice as large as that for high-school students.

For comparison, I estimate the tax elasticities from the standard cross-sectional model, which does not include peer effects, for the same tax increases. The results are shown in the second panel of Table 12.[45] The following findings are highlighted. First, the estimated elasticities from the standard model are consistently larger than those from the social interaction model. As stated earlier (see section 3.3), the peer effects are

45. The estimated parameters from the standard model are available from the supplementary web site. Since no peer interactions are assumed in the standard model, the parameter estimates can be obtained by standard logistic regression.

identified from excessive cross-group variations in choices that cannot be explained by differences in fundamentals. Thus, the standard model overstates the impact of tax on smoking decisions, which in practice, may be amplified by peer interactions. Second, the tax elasticities for middle-school students are significantly larger than those for high-school students, based on models with and without peer effects. Third, however, the difference in tax elasticities between middle-school and high-school students is much greater according to the social interaction model than according to the standard model. This result indicates that the larger difference in tax elasticities is due to the different magnitudes of peer effects between middle- and high-school students. Since middle-school students have larger peer effects than high-school students, the tax effects on middle school students might be greater than those on high-school students.

Do the estimated tax elasticities explain changes in the smoking rate over time? I examine this issue by computing out-of-sample predictions. Youth smoking rates have been falling steadily since the early 2000s. According to the data from the Monitoring the Future Survey, between 2000 and 2004, the smoking rate among 8th graders fell from 14.6 percent to 19.2 percent, and the smoking rate among 12th graders decreased from 31.4 percent to 25.0 percent. During the same period, federal and state excise taxes increased from 75.9 cents to 111.6 cents per pack.[46] One would expect at least part of the decline in youth smoking rates to have been due to these tax increases. Based on the estimated elasticities from the social interaction model above, [47] the tax increase can explain 31.3 percent of the 5.4 percentage-point decrease in the smoking rate of 8th graders, and can explain 27.3 percent of the 6.4 percentage-point decrease in the smoking rate among 12th graders. This suggests that the tax increase can explain about one third of the smoking trends of middle- and high-school students. Thus, although the estimated tax elasticities appear small, they can explain a substantial part of the time-series changes since 2000. According to a similar estimate reported by Gruber and Zinman (2000) and Gruber (2001), tax increases explain about 26 percent of the the changes in smoking rates between 1991 and 1997 among 12th-grade, which is comparable with the finding of this paper. Note that the explanatory power of the tax increases is similar for middle- and high school students.

The explanatory power of the estimated tax elasticities from the standard model is different. On the basis of the results reported in the second panel of Table 12, the tax shift explains 33.3 percent of the time-series variation in the smoking trend among 8th graders and explains 44.5 percent of the time-series variation in smoking by 12th-grade students between 2000 and 2004.[48] The finding that tax has a greater effect on high-school students than on middle-school students is unconvincing, given that the smoking trends of middle-school and high-school students are similar. However, there is no strong evidence linking reduced smoking rate among middle-school students to other changes, either in background characteristics or in regulatory policies.

As stated earlier, the tax effects are magnified by multiplier effects. Table 13 reports the decomposition into these direct and indirect effects of tax increase on smoking prevalence for all students from grades six to 12. The first and second columns show

46. To be more precise, between 2000 and 2004, the average state excise tax increased from 41.9 cents to 72.6 cents, and the federal excise tax increased from 34.0 cents to 39.0 cents.

47. Assuming a tax change from 2000 to 2004 of about 40 cents, I used the estimated elasticity of -0.247 for middle-school students in the 8th grade, and the estimated elasticity of -0.119 for high-school students in the 12th grade.

48. I used the estimated elasticity -0.262 for 8th-grade students and the estimated elasticity -0.193 for 12th grade students.

TABLE 13

Decomposition of Tax Effects on Smoking Prevalence

tax increases	total effect	direct effect	social multiplier
20 cents	-0.857	-0.533	1.610
40 cents	-1.677	-1.055	1.590
60 cents	-2.458	-1.567	1.569
80 cents	-3.203	-2.068	1.549
100 cents	-3.917	-2.559	1.531
120 cents	-4.598	-3.039	1.513

Note- Data are composed of all students between the grades of 6 and 12 of the four major races in the United States (whites, blacks, Hispanics and Asians). Data are restricted to students in the schools for which the 2000 NYTS collects at least ten male and female students. The sample size is $N = 29385$ from $S = 305$ schools. All simulated values are computed from the average over 1000 experiments. Social multiplier is defined by the ratio of total effect to direct effect.

the total effect and the direct effect, respectively. The third column shows the implied social multiplier effects, which are the factors by which the externality raises the direct effect through peer interactions. As shown, the multipliers range from 1.513 to 1.612 for the tax increases examined. This implies that peer effects raise the direct tax effect on youth smoking by a factor of more than 1.5. This evidence of significant multiplier effects on the tax elasticities supports the hypothesis that peer interactions strongly affect youth smoking behavior. [49]

In summary, there is evidence to suggest that cigarette taxes can substantially reduce youth smoking. Tax increases explain about one third of the observed recent reductions in the smoking rates of both middle- and high-school students. It should be stressed, however, that tax is not the only effective tool for reducing youth smoking. For example, *Healthy People 2000* sets the objective of more than doubling the average federal and state taxes on cigarettes to \$2.00 per pack. According to the results in Table 12, a tax increase of 120 cents per pack (a cigarette tax of \$2.10 per pack) could reduce the smoking rates of middle-school and high-school students by 4.6 percentage points, from 18.9 percent to 14.3 percent. While this tax-induced reduction is significant, it is far short of the target set by *Healthy People 2000*, which is to reduce youth smoking to half of its current rate.

7. 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-school and high-school students to examine the hypothesis that peer interactions play an important role in explaining variations in youth smoking behavior.

49. Note that the estimated social multiplier effects decrease systematically as the hypothetical tax increases. The intuitive explanation is as follows. Suppose that tax increases. The social interaction model implies that the latent smoking utility of a person , which is assumed to be linear in background characteristics, decreases *proportionally* to the tax rise, due to a direct tax effect. At the same time, the tax rise reduces the fraction of smokers in the person's peer group, so that the utility of the person decreases further through the indirect effect of the reduced fraction of smokers. However, the decrease in utility due to the indirect effect is *not proportional* to the tax increase because the fraction of smokers is bounded between 0 and 1. The larger the tax change, the closer is the fraction of smokers to the upper or lower bound. Therefore, as the tax increases, the direct tax effect dominates the indirect tax effect, and the social multiplier diminishes.

The model presented in this paper specifies how each person's smoking behavior is related to the smoking behavior of peers through utility. The behavioral model incorporates a utility-maximization framework by 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 by 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 conclude by emphasizing that the empirical analysis of this paper is merely the first step of a wider study into peer effects on smoking decisions. To develop a comprehensive policy strategy to reduce the prevalence of tobacco smoking, further research should be implemented to clarify the remaining issues. In particular, the model did not incorporate an addiction mechanism where reinforcing feedback operates across past-self, present-self, and future-self within an individual. Instead, the model of this paper focuses on peer interactions where a reinforcing feedback loop is only at work across different individuals. Undoubtedly, it would be worthwhile to analyze both dimensions of reinforcing feedbacks. This analysis would require a rational expectations model where young people optimally choose their current smoking decision based on past smoking choices by both themselves and their peers, as well as on their anticipation of the future smoking choices of both themselves and their peers. The dynamic analysis featuring forward-looking expectations is a fertile ground for future research.

APPENDIX A. PROOFS AND DERIVATIONS OF EQUATIONS

Appendix A.1. *Proof of Theorem 1*

I show that the conditional probabilities π_i are compatible under the steady state distribution P^*. I will use the phrase *compatible* to refer to the case in which a set of conditional probabilities are generated by a common joint distribution (see Arnold et al. (1999)).

Consider the conditional probability of P^* on ω_i given ω_{-i} and given $\mathbf{x}$. A straightforward computation yields

$$P^*(\omega_i|\omega_{-i};\mathbf{x}) = \frac{P^*(\boldsymbol{\omega}|\mathbf{x})}{P^*(\boldsymbol{\omega}_+|\mathbf{x}) + P^*(\boldsymbol{\omega}_-|\mathbf{x})} \tag{A21}$$

where $\boldsymbol{\omega} = (\omega_i, \omega_{-i})$,$\boldsymbol{\omega}_+ = (+1, \omega_{-i})$, and $\boldsymbol{\omega}_- = (-1, \omega_{-i})$. The denominator is the marginal distribution $P^*(\omega_{-i}|\mathbf{x}) = P^*(\boldsymbol{\omega}_+|\mathbf{x}) + P^*(\boldsymbol{\omega}_-|\mathbf{x})$. For convenience, I rewrite Equation (A21) by dividing the both denominator and numerator by $P^*(\boldsymbol{\omega})$ as

$$P^*(\omega_i|\omega_{-i};\mathbf{x}) = \begin{cases} [1 + P^*(\boldsymbol{\omega}_-|\mathbf{x})/P^*(\boldsymbol{\omega}_+|\mathbf{x})]^{-1} & \text{if} \quad \omega_i = +1, \\ [1 + P^*(\boldsymbol{\omega}_+|\mathbf{x})/P^*(\boldsymbol{\omega}_-|\mathbf{x})]^{-1} & \text{if} \quad \omega_i = -1. \end{cases} \tag{A22}$$

Suppose that $\omega_i = +1$. Then I compute the value $P^*(\boldsymbol{\omega}_-|\mathbf{x})/P^*(\boldsymbol{\omega}_+|\mathbf{x})$ in Equation (A22) as

$$P^*(\boldsymbol{\omega}_-|\mathbf{x})/P^*(\boldsymbol{\omega}_+|\mathbf{x}) = \exp -\{Q(\boldsymbol{\omega}_+|\mathbf{x}) - Q(\boldsymbol{\omega}_-|\mathbf{x})\}, \tag{A23}$$

using the definition of P^* in Equation (2.6). By substituting the Qs by Equation (2.7), I obtain the following result

$$P^*(\omega_i|\omega_{-i};\mathbf{x}) = \Big[1+\exp -\big\{b_i(\mathbf{x}_i)+\textstyle\sum_{j\neq i}\rho_{ij}\omega_j\big\}\Big]^{-1} \tag{A24}$$

A similar result follows when $\omega_i = -1$. Under the assumption 1 and 2, this is equal to $\pi_i(\omega_i|\omega_{-i},\mathbf{x}_i)$, which is the conditional probability of smoking decision for person i. Thus, I proved that the set of conditional probability π_i are compatible under the distribution P^*.

The same induction steps as in Amemiya (1975) prove that the P^* is a unique joint distribution from which the set of distribution π_i are generated from.

I then turn attention to the convergence of the interaction process $[\mathbf{y}(0),\mathbf{y}(1),\cdots,\mathbf{y}(t)]$. The transition of the stochastic process is given by rule (2.4) with the updating probability π_i. This type of the Markov chain is generally referred as *Glauber* dynamics, or *Gibbs sampler* dynamics. According to the theory of Markov chain (see Geman and Geman 1984; Guyon 1995), if every N person in a sequence $(d(0),d(1),d(2),\cdots)$ is chosen infinitely often, the Markov chain $[\mathbf{y}(0),\mathbf{y}(1),\cdots,\mathbf{y}(t)]$ converges on a common joint distribution under which the set of conditional probabilities π_i are compatible, which turn out to be the distribution P^* from the discussion above. This result is independent of the initial condition $\boldsymbol{\omega}(0)$, and the result of Eq (2.5) follows. The only necessary condition for applying the convergence result follows from the assumption 3. ‖

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

I begin with notations. Given the vector of individual background 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.

By using the new notations, the Q-function in Equation (5.17) can be rewritten as

$$\begin{aligned} Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta}) &= \frac{1}{2}\sum_i \omega_i(\delta_\ell+\mathbf{x}_{1i}\boldsymbol{\beta}_1+\mathbf{x}_{2i}\boldsymbol{\beta}_2)+\frac{1}{4}\sum_i\sum_j \rho_{ij}\omega_i\omega_j \\ &= \left(\frac{1}{2}\sum_i(\delta_\ell+\mathbf{x}_{2i}\boldsymbol{\beta}_2)\tau_\ell\right)+\left(\frac{1}{2}\sum_i \omega_i\mathbf{x}_{1i}\boldsymbol{\beta}_1+\frac{1}{4}\sum_i\sum_j\rho_{ij}\omega_i\omega_j\right). \end{aligned} \tag{A25}$$

The last equation follows from the restriction $\tau_\ell=\sum_i\omega_i$. I use the following notation for convenience: $C_\ell=\frac{1}{2}\sum_i(\delta_\ell+\mathbf{x}_{2i}\boldsymbol{\beta}_2)\tau_\ell=\frac{NL}{2}(\delta_\ell+\mathbf{x}_{2i}\boldsymbol{\beta}_2)\tau_\ell$. Then, the Q function is

$$Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta}) = C_\ell + Q_1(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta}). \tag{A26}$$

Bayes rule implies that the conditional probability of P^* on $\boldsymbol{\omega}_\ell$ given τ_ℓ is computed as follows

$$\begin{aligned} P^*(\mathbf{y}_\ell=\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta};\tau_\ell) &= \frac{\exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}_\ell\in\Omega_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})}\Bigg/\frac{\sum_{\boldsymbol{\omega}_\ell\in B_\ell}\exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}_\ell\in\Omega_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})} \\ &= \frac{\exp Q(\boldsymbol{\omega}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})}{\sum_{\boldsymbol{\eta}_\ell\in B_\ell}\exp Q(\boldsymbol{\eta}_\ell|\mathbf{x}_\ell,\boldsymbol{\theta})}. \end{aligned} \tag{A27}$$

Plug Equation (A26) into Equation (A27). Since the $\exp C_\ell$ is constant for any profile that belong to B_ℓ, it is canceled out from the numerator and the denominator above. Thus Equation (5.17) follows. ‖

APPENDIX B. AN EXAMPLE OF A MARKOV CHAIN

Consider a Markov chain with following transition: Suppose that pair of persons i and j are randomly selected. Let $\sigma(\boldsymbol{\omega})=(\sigma_1,\cdots,\sigma_N)$ be a vector that swaps i and j elements of $\boldsymbol{\omega}$. In other words, $\sigma_i=\omega_j$ and $\sigma_j=\omega_i$. Consider the following transition of profile $\mathbf{y}(t+1)$ from $\mathbf{y}(t)=\boldsymbol{\omega}\in B$ such as

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

where I define

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

The transition states that the candidate value in next period is proposed by swapping choices between the randomly selected persons i and j, and is then accepted or rejected according to the Q-function at the candidate value relative to the Q-function at the current value. Such an algorithm is often referred to as the *Metropolis* sampling algorithm in the statistical literature. Note that whenever $y(t) \in B$ then $y(t+1) \in B$. That is, the Markov chain proposed above satisfies the summation restriction imposed by B where $\sum_i \omega_i = \tau$ for $\boldsymbol{\omega} \in B$. Since two elements of $\boldsymbol{\omega}$ are just swapped, it follows that $\sum_i \sigma_i = \sum_i \omega_i = \tau$, and thus $\sigma(\omega) \in B$.

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. The reason is as follows: 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.

Acknowledgment. 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, Partha Deb, Nobuyuki Hanaki, Daiji Kawaguchi, Edward Norton, Tsunao Okumura and Fumio Ohtake. I also appreciate the editor, Bernard Salanie, and two anonymous referees who provided useful comments on the previous draft. Of course, the responsibility of all remaining errors is mine. I am grateful to the Osaka University Institute of Social and Economic Research and Duke University the Center for Child and Family Policy for their hospitality at various stages of this project. The paper received the Osaka University Institute of Social and Economic Research/Moriguchi Prize.

REFERENCES

Amemiya, T. (1975). Qualitative response models. *Annals of Economics and Social Measurement 4*, 363–372.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of The Royal Statistical Society, Series B 82*, 283–301.

Arcidiacono, P. and S. Nicholson (2005). Peer effects in medical school. *Journal of Public Economics 89*, 327–350.

Arnold, B. C., E. Castillo, and J. M. Sarabia (1999). *Conditional Specification of Statistical Models*. Springer Series in Statistics. Springer Verlag.

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

Bertrand, M., E. F. P. Luttmer, and S. Mullainathan (2000). Network effects and welfare cultures. *Quarterly Journal of Economics 115*, 1019–1055.

Bisin, A., U. Horst, and O. Özgür (2006). Rational expectations equilibria of economies with local interactions. *Journal of Economic Theory 127*, 74–116.

Blume, L. E. (1993). The statisticalmechanics of strategic interaction. *Games and Economic Behavior 5*, 387–424.

Blume, L. E. (2003). How noise matters. *Games and Economic Behavior 44*, 251–271.

Blume, L. E. and S. N. Durlauf (2003). Equilibrium concepts for social interaction models. *International Game Theory Review 5*, 193–209.

Bresnahan, T. F. and P. C. Reiss (1990). Entry in monopoly markets. *Review of Economic Studies 57*, 531–553.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*, 235–260.

Case, A. C. and L. F. Katz (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. (National Bureau of Economic Research, Inc, NBER Working Papers: 3705).

Center for Disease Control and Prevention (2001). Youth tobacco surveillance united states - 2000. *CDC Surveillance Summaries MMWR 50.*

Chaloupka, F. J. (1991). Rational addictive behavior and cigarette smoking. *Journal of Poitical Economy 99*, 722–742.

Chaloupka, F. J. and M. Grossman (1996). Price, tobacco control policies and youth smoking. (National Bureau of Economic Research, Inc, NBER Working Papers: 5740).

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

Conley, T. G. and G. Topa (2003). Identification of local interaction models with imperfect location data. *Journal of Applied Econometrics 18*, 605–618.

DeCicca, P., D. Kenkel, and A. Mathios (2000). Racial difference in the determinants of smoking onset. *Journal of Risk and Uncertainty 21*, 311–340.

Durlauf, S. N. and J. R. Walker (2001). Social interactions and fertility transitions. In J. Casterline (Ed.), *Diffusion Processes and Fertility Transition: Selected Perspectives*, pp. 115–137. Washington, D.C.: National Academy Press.

Ellisson, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*, 1047–1071.

Emery, S., M. M. White, and J. Pierce (2001). Does cigarette price influence adolescent experimentation. *Journal of Health Economics 20*, 261–270.

Flay, B. R., J. R. d'Avernas, M. W. Kersell, and K. B. Ryan (1983). Cigarette smoking: Why young people do it and ways of preventing it. In J. M. Patrick and F. Phillip (Eds.), *Pediatric and Adolescent Behavioral Medicine*, pp. 132–181. New York: Springer.

Föllmer, H. (1974). Random economics with many interaction agents. *Journal of Mathematical Economics 1*, 51–62.

Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics 83*, 257–268.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 721–741.

Geyer, C. J. (1991). Markov chain monte carlo maximum likelihood. In *23rd Symposium on Interface, Computer Science and Statistics*, pp. 156–163.

Geyer, C. J. and E. A. Thompson (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of The Royal Statistical Society, Series B 54*, 657–699.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *Quarterly Journal of Economics 111*, 507–548.

Glaeser, E. L. and J. A. Scheinkman (2000). Non-market interactions. (National Bureau of Economic Research, Inc, NBER Working Papers: 8053).

Glaeser, E. L. and J. A. Scheinkman (2001). Measuring social interactions. In S. N. Durlauf and P. H. Young (Eds.), *Social Dynamics*, pp. 83–132. Cambridge: MIT Press.

Gruber, J. (2001). Youth smoking in the 1990's: Why did it rise and what are the longrun implications? *American Economic Review 91*, 85–90.

Gruber, J. and J. Zinman (2000). Youth smoking in the us: Evidence and implications. (National Bureau of Economic Research, Inc, NBER Working Papers: 7780).

Guyon, X. (1995). *Random Fields on a Network Modeling, Statistics and Applications.* New York: Springer-Verlag.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*, 931–959.

Huffer, F. W. and H. L. Wu (1998). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. *Biometrics 54*, 509–524.

Johnston, L. D., P. M. O'Malley, and J. G. Bachman (2004). Monitoring the future national survey results on drug use, 1975-2003. volume i: Secondary students. (National Institute on Drug Abuse, NIH Publication No. 04-5507).

Lehmann, E. L. and G. Casella (1998). *Theory of point estimation.* New York: Springer.

Liebenstein, H. (1950). Bandwagon, snob, and veblen effects in the theory of consumer demand. *Quarterly Journal of Economics 64*, 183–207.

Manski, C. F. (1993). Identification of endogenous social effects the reflection problem. *Review of Economic Studies 60*, 531–542.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophyly in social networks. *Annual Review of Sociology 27*, 415–444.

Norton, E. C., R. C. Lindrooth, and S. T. Ennett (1998). Controlling for the endogeneity of peer substance sse on adolescent alcohol and tobacco use. *Health Economics 7*, 439–453.

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

Powell, L. M., J. A. Tauras, and H. Ross (2005). The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of Health Economics 24*, 950–968.

Shrum, W., N. H. Cheek, and S. M. D. Hunter (1988). Friendship in school: Gender and racial homophyly. *Sociology of Education 61*, 227–239.

Tamer, E. (2003). Incomplete simultaneous discrete response: Model with multiple equilibria. *Review of Economic Studies 70*, 147–165.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies 68*, 261–295.

U.S. Department of Health Human Services (1998). *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.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Weinberg, B. A., P. B. Reagan, and J. J. Yankow (2004). Do neighborhoods affect hours worked? evidence from longitudinal data. *Journal of LaborEconomics 22*, 891–924.

Young, P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.