Peer effects in sexual initiation: Separating demand and supply mechanisms

Seth Richards-Shubik
Department of Economics, Lehigh University

Most work on social interactions studies a single, composite effect of interactions within a group. Yet in the case of sexual initiation, there are two distinct social mechanisms—peer-group norms and partner availability—with separate effects and different potential interventions. Here I develop an equilibrium search and matching model for first sexual partners that specifies distinct roles for these two mechanisms as part of demand and supply. I estimate the model using a national sample of high school students, with data over time on individual virginity status. The results indicate that peer-group norms have a large effect on the timing of sexual initiation for both boys and girls. Changes in opposite-gender search behavior (i.e., partner availability) also have a large impact on initiation rates for boys, but not for girls.

Keywords. Social interaction models, mechanisms, sexual activity, youth, structural estimation.

JEL classification. C31, C33, J13.

1. Introduction

About one-half of the students in grades 9–12 in the United States are sexually experienced (CDC (2008)). Sexual debut is a normal part of human development, but early onset and high prevalence of sexual activity among adolescents raises concerns, largely because of associated risks such as unplanned pregnancy and sexually transmitted dis-

eases. Additionally, as with many risky behaviors, peers are often pointed to as a major influence on the decisions adolescents make about sex.

A large body of research in economics examines peer effects, but the ability of this work to inform policy is often limited to one class of interventions: regrouping individuals. The models considered in the empirical literature typically define an endogenous social effect as the change in the probability of an outcome for an individual caused by a change in the distribution of that outcome within some reference group. This represents a *composite* effect of social interactions, which may combine the effects of multiple, distinct mechanisms. The composite effect can be sufficient to determine the impact of regrouping policies, which are important in contexts such as education (e.g., ability tracking, desegregation busing). However for interventions that target particular mechanisms behind an endogenous social effect, assessing the separate effects of these mechanisms is crucial for making policy recommendations. This is especially relevant for school-based interventions related to sexual activity because there are (at least) two plausible social mechanisms at play: social norms among peers and the availability of partners at school. These mechanisms can be well understood as aspects of demand and supply in a market for sexual partners. Moreover, they are targeted by different interventions. Most interventions that address the effect of peer norms on demand use a direct educational approach. One example is the "Safer Choices" program, which consists of a number of classroom sessions, some of which are devoted to social norms, such as the following session:

> *The Safest Choice: Deciding Not to Have Sex.* Students learn about "social norms." They discuss perceptions of how many of their peers have had sex and how these perceptions compare to actual statistics. Using role-playing, students also learn refusal skills (Manlove, Romano-Papillo, and Ikramullah (2004, p. 31)).

On the supply side, the most obvious policy options to restrict the availability of partners involve some form of segregation, whether by gender or age. Single-sex schools represent one way to do this, and about 2% of U.S. high school students attend gender-segregated schools.[1] A less drastic option is to isolate the ninth grade from the older grades in high school, as in school districts where the ninth grade is in the middle school.

In this paper, I estimate an equilibrium model of the market for sexual partners in high school, so as to measure the separate effects of peer norms and partner availability on sexual initiation, and thereby to inform our views on the potential effectiveness of interventions such as these. In the model, the demand from each individual depends on the expected costs and benefits of sex, which is influenced by the share of same-gender peers who are nonvirgins. This is the effect of peer norms. The model uses a search and matching framework in which individual demand appears as the decision to search for a sexual partner. The probability of finding a partner depends on the search decisions of others in the market, which determines the arrival rate of match offers. Accordingly the effect of partner availability can be defined as the change in the probability of finding a match due to changes in the search behavior among others at school.

---

[1]Calculated with data from the National Center for Education Statistics Public Elementary/Secondary School Universe Survey and Private School Universe Survey for the 2009–2010 academic year. Downloaded from http://nces.ed.gov/ccd/elsi/ on 2/11/2014.

The model is dynamic, as in the job search literature (e.g., Eckstein and Wolpin (1990)), which is unlike the static matching models in other recent work on marriage and dating markets (e.g., Choo and Siow (2006), Arcidiacono, Beauchamp, and McElroy (forthcoming)). To solve the model, I follow the literature on dynamic discrete games (Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007)) and use observable equilibrium state transition probabilities as rational beliefs.[2] This approach greatly reduces the computational burden of estimation.

To estimate the model, I use longitudinal data on individual virginity status from the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative but highly clustered sample of U.S. high school students in the mid-1990s. The observation of outcomes over time allows me to use the initial nonvirginity rates when a cohort enters high school to control for common unobserved factors, which would otherwise bias the estimates of social effects. I specify a distribution of permanent preferences toward sex that is a function of these initial rates and thus is naturally correlated within schools. This captures unobserved factors that influence the demand for sex that are time-invariant and pertain to the community where a school is located. I show that this approach is equivalent for identification to including school fixed effects in the distribution of preferences, under reasonable assumptions. The variation that identifies the effects of peer norms and partner availability under this approach then comes from unpredicted, random innovations in group nonvirginity rates, conditional on their initial rates.

In addition to this issue of potentially correlated unobservables, other challenges for the identification of social interactions models include selection into peer groups (Moffitt (2001), Brock and Durlauf (2001b)) and the "reflection" problem (Manski (1993)). To avoid the selection problem, I use gender–grade cohorts to define peer groups, rather than endogenous social groups like sports teams or nominated friends. The selection of students into schools is another concern, but this would be addressed with the initial nonvirginity rates as part of the more general problem of correlated unobservables within schools. The reflection problem does not apply in my model for two reasons. First, the model is nonlinear and, as Brock and Durlauf (2001b) show, the reflection problem depends on linearity. Second, and more distinctively, the standard assumption of stable preferences over time provides a key identifying restriction. A final challenge for the identification of my model is the separation of the arrival rate from the search probability. To address this, I use data on the arrival of subsequent partners after the first, which nonparametrically identifies the arrival rate.

The results I obtain indicate that within-gender peer norms have a large effect on the timing of sexual initiation for both boys and girls. In a counterfactual simulation that removes the peer influence on search decisions, the number of individuals who initiate sex during high school falls by 26% for boys and 20% for girls. On the other hand, changes in the availability of partners at school appear to have a large impact on the timing of initiation for boys, but not for girls. In a simulation that removes the effect of partner availability at school, the number of boys who initiate sex during high school

---

[2]These papers build on the method originally developed to solve individual dynamic models by Hotz and Miller (1993). However, I do not use observed individual choice probabilities, only aggregate state transitions.

falls by 37% while the number of girls falls by only 12%. This is possible because the model includes an external market for partners that appears as an exogenous component of the arrival rate. This component varies with gender and grade, and the estimates indicate that girls have a substantially higher rate of match offers from the external market, which makes them less sensitive to changes in the availability of boys in their local school market. Finally, I simulate the impact of isolating the ninth grade from the older grades in high school to restrict the supply of partners. The results indicate only a small impact on sexual initiation in the ninth grade and are not statistically significant.

The existing literature does not provide any results that are directly comparable to these estimates. However, two recent papers are sufficiently similar that some loose comparisons are possible. Card and Giuliano (2013) estimate the peer influence of same-gender best friends on sexual initiation, with a selection model to control for homophily. Arcidiacono, Beauchamp, and McElroy (forthcoming) estimate a static matching model of high school romantic relationships with and without sexual relations. In both cases, the magnitudes of the social effects they recover are roughly similar to the effects of the relevant mechanism in my model (see Section 5 for a discussion). More broadly, several other studies have found large composite effects of social interactions among adolescents on various risky behaviors, including sexual initiation (Fletcher (2007)) and teenage childbearing (Case and Katz (1991), Evans, Oates, and Schwab (1992)), as well as criminal activity, high school completion, substance abuse, and obesity (Case and Katz (1991), Gaviria and Raphael (2001), Lundborg (2006), Clark and Lohéac (2007), Trogdon, Nonnemaker, and Pais (2008)).

To provide further support for my structural estimates and intuition about their identification, I estimate a series of simple hazard models for sexual initiation that include a composite effect of social interactions. These composite effects come from lagged nonvirginity rates while the initial rates are used as controls, so they reflect essentially the same variation in the data that identifies the social effects in my model. The estimates of these composite effects are robust across a variety of specifications. In addition these hazard models demonstrate that the initial nonvirginity rates absorb substantial heterogeneity in unobserved factors that otherwise would bias estimates of social effects.

A small number of authors have previously considered social interactions in duration models and other nonlinear dynamic models. Brock and Durlauf (2001b) discuss the identification of parametric, continuous-time duration models, and Sirakaya (2006) estimates such a model applied to recidivism. de Paula (2009) develops a nonparametric test for social interactions in duration models based on simultaneous exits. Nakajima (2007) estimates a dynamic binary choice model of smoking behavior with peer effects.[3]

These papers and most other empirical studies of social interactions recover some kind of composite effect, which could reflect multiple mechanisms. However, the literature recognizes an important distinction between the composite effects of endogenous versus exogenous social interactions.[4] As Manski (1993) notes, only endogenous

---

[3]In comparison to the model developed here, the agents in Nakajima (2007) are myopic and move in an exogenous sequence. The steady-state distribution from that model was matched to cross-sectional data.

[4]Exogenous social effects arise from the distribution of exogenous characteristics within the reference group rather than the distribution of endogenous outcomes.

social effects generate multipliers on individual-based interventions. In addition, many empirical papers define multiple reference groups based on exogenous characteristics (e.g., race, gender) and recover separate endogenous social effects from each group (e.g., Nakajima (2007)). Differences in the effects from different reference groups can be interpreted as evidence that some mechanisms are more important than others.[5] For example, Mas and Moretti (2009) find that productivity spillovers among supermarket cashiers are stronger from co-workers positioned behind a given individual rather than in front of her, and from this they infer that the dominant mechanism is one of social pressure rather than prosocial behavior. A handful of papers go further and formally derive testable implications of specific mechanisms for endogenous social interactions, which they then compare with patterns in the data. Drewianka (2003) considers endogenous social effects in the marriage market and evaluates certain implications of a mechanism related to market conditions rather than preference interactions. De Giorgi and Pellizzari (2014) derive distinct implications from three alternative mechanisms for endogenous social effects in academic performance, and they find evidence that supports a particular mechanism of mutual insurance.[6] Separately, Fruehwirth (2014) shows that predictions for regrouping policies may be incorrect if social effects involve unobserved endogenous interactions while the estimated effects do not account for this underlying structure. Relative to these diverse efforts to examine mechanisms in social interactions, to the best of my knowledge, the work presented here is the first to directly recover the separate effects of multiple mechanisms behind an endogenous social effect.

The paper proceeds with the model in the next section. Section 3 then describes the data and presents estimates of the composite effects of social interactions. Section 4 describes the estimation procedure and contains detailed arguments on identification. Section 5 presents the results and counterfactual simulations. Section 6 concludes. Appendices and replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/249/supplement.pdf and http://qeconomics.org/supp/249/code_and_data.zip.

2. A SEARCH AND MATCHING MODEL FOR FIRST SEX

The model describes a discrete-time dynamic process leading to sexual initiation. Each period, virgins decide whether or not to search for their first sexual partners. For those who search, the probability of finding a partner depends on the search behavior of others within their local market. This local market is defined within the student body at a high school. There is also an external market for partners, which appears as an exogenous probability of finding a partner from outside the school.

---

[5]As Blume, Brock, Durlauf, and Ioannides (2011) point out, however, there is a distinction between effects defined in terms of variables rather than economic mechanisms. They write, "In the econometrics literature, contextual and endogenous social interactions are defined in terms of types of variables rather than via particular mechanisms. This can delimit the utility of the models we have, for example, if the particular mechanisms have different policy implications." (p. 941)

[6]In a similar vein, Duflo, Dupas, and Kremer (2011) derive and evaluate testable implications of multiple mechansms in a model with exogenous social effects.

The model abstracts from certain aspects of adolescent sexual behavior that would add complications without greatly enhancing the analysis of social influences in sexual initiation per se. First, there is no constraint on the number of partners per period. Although a single partner per period is the most common, multiple partners (observed as overlapping relationships) also appear in the data. To incorporate this distinction in the model, I would need to specify multiple types of relationships (exclusive and nonexclusive) and include a dissolution rate for exclusive relationships. Then the arrival rate of partners would depend in part on the share of exclusive relationships, and agents would need to keep track of this aspect of the market, which would greatly expand the state space.

Second, payoffs relate directly to virginity status. All the costs and benefits of sexual activity, such as the risk of pregnancy or the frequency of sex, are embedded in the expected utility of nonvirginity. Accordingly, any subsequent decisions related to sexual activity are suppressed (e.g., contraceptive use).[7] Further decisions and additional structure in the payoffs are not needed for this analysis because, for a virgin, it is the overall expected utility of novirginity that determines whether he or she wants to search for a partner. Third, nonvirgins are assumed to stay in the market and continually search for new partners. This allows individuals to have more than one partner during high school, which is true for a substantial portion of the population, without further complicating the model.
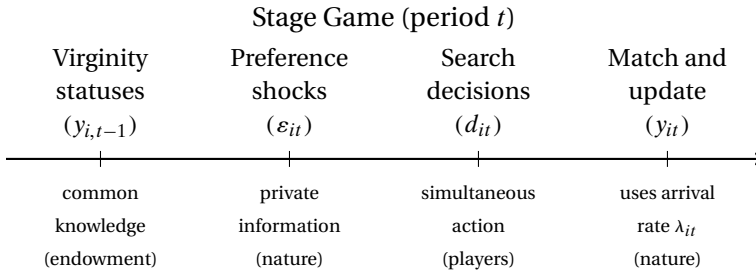
Finally, match probabilities do not depend on own or partner characteristics. Including them would introduce sorting behavior, which is not the focus of this paper. Consequently, the arrival rate in the model averages over any individual heterogeneity and any differences related to the characteristics of opposite-gender searchers. To the extent that arrival rates are in fact heterogeneous, the model misassigns the effect of such characteristics to the search decision. However, the characteristics one would think to use to add heterogeneity to the match probabilities are typically permanent attributes. In contrast, the primary objects of interest—the effects of peer norms and partner availability—are identified from changes in nonvirginity rates over time, not permanent attributes (see Section 4.2).

2.1 Model specification

The model is based on a repeated game of incomplete information, where each period (t) corresponds to a stage game. The virginity statuses from the previous period among the individuals (i) in the local market are denoted $y_{i,t-1} \in \{0, 1\}$, with 0 meaning virginity. These statuses are common knowledge, as are the individuals' gender, age, and permanent preference characteristics (described below). In each stage, nature first draws preference shocks for each individual, ε_{it}, which are private information. Then the virgins in the market simultaneously make their search decisions, denoted $d_{it} \in \{0, 1\}$. Nonvirgins always search. The arrival rates faced by each individual, λ_{it}, are determined endogenously by these joint search decisions. Then nature randomly assigns some searchers

---

[7]Also, there is no decision to accept a match offer. This is not needed because all matches produce the same payoff for an individual.

to find a partner based on these probabilities, and virgin searchers who find a partner transition to become nonvirgins. These steps are summarized in the diagram below.

Stage Game (period t)

Virginity statuses $(y_{i,t-1})$	Preference shocks (ε_{it})	Search decisions (d_{it})	Match and update (y_{it})
common knowledge (endowment)	private information (nature)	simultaneous action (players)	uses arrival rate λ_{it} (nature)

Gender is binary, indicated with b for boys and g for girls.[8] All functions and parameters are gender-specific, but gender subscripts are generally suppressed unless needed for clarity. Age, a, is defined socially as the quarter within grade in high school. The model starts with the fall of ninth grade ($a = 1$) and ends with the spring of twelfth grade ($a = 15 \equiv A$). Time is also measured in quarters, and is needed separately from age to track multiple cohorts at once. However, in the exposition, the model is typically presented from the perspective of a reference cohort for which time and age are equal ($a_{it} = t$).

The arrival rate, λ_{it}, gives the probability of finding a partner in the current period. In the main specification it is a function of the proportion of searchers among the opposite gender at the school. This proportion is denoted N_{it}, and it includes both virgins and nonvirgins. The function to yield λ_{it} from this proportion is specified with the logistic cumulative distribution function (CDF)

$$\lambda_{it} = \lambda_{a_{it}}(N_{it}) \equiv \frac{\exp(\lambda_{0a_{it}} + \lambda_1 N_{it})}{1 + \exp(\lambda_{0a_{it}} + \lambda_1 N_{it})}. \tag{1}$$

The arrival rate is positive even if there are zero searchers at a school, which reflects the existence of an external market for partners. The parameters λ_{0a} vary with age to allow for changes in the amount of contact with the external market as students progress to older grades. However, the main parameter of interest in (1) is λ_1, which gives the effect of partner availability at school (i.e., in the local market) on the arrival rate.

This specification of the arrival rate does not depend on the search behavior by people of the same gender as the individual, meaning it essentially ignores competition for partners. The absence of competition can be justified by the lack of a constraint on the number of partners, so that matches need not be one-to-one within each time period. However, I also consider an alternative version of the matching technology that makes the opposite assumption. The alternative specification of the arrival rate uses the ratio of the numbers of opposite-gender to own-gender searchers: for example, $\lambda_{it} = \lambda_{a_{it}}(N_{it}^g / N_{it}^b)$ for a boy, where N^g and N^b here denote the *numbers* rather than proportions of searchers of each gender. This mimics the ratio of job seekers to vacancies

---

[8]The model pertains to heterosexual sex, so a partner must be of the opposite gender.

that is commonly used in the macroeconomic labor literature such as Mortensen and Pissarides (1994). To motivate this kind of matching function, it is typically assumed that matches are strictly one-to-one per period.[9]

Individuals derive utility from their virginity status. The per-period payoff for being sexually experienced is a linear combination of age, the proportion of peers who are already nonvirgins, denoted $Y_{i,t-1}$, a permanent individual component, ω_i, and an independent and identically distributed (IID) mean-zero preference shock, ε_{it}. Peers are individuals of *the same gender in the same grade* as individual i. The per-period utility for a nonvirgin is thus

$$u(a_{it}, Y_{i,t-1}, \omega_i, \varepsilon_{it}) \equiv \overbrace{\alpha a_{it} + \gamma Y_{i,t-1}}^{\hat{u}_{it}} + \omega_i + \varepsilon_{it}. \tag{2}$$

The per-period utility for a virgin is normalized to zero.

The term $\gamma Y_{i,t-1}$ represents the effect of peer norms. This is a standard specification for a social component of utility, as in Brock and Durlauf (2001a).[10] To be precise about the interpretation, the social utility term generically expresses the effect of lagged peer nonvirginity rates on the flow utility of virginity status. I interpret this as an effect of social norms based on research on adolescent sexual behavior in the sociology and psychology literatures (e.g., Kinsman, Romer, Furstenberg, and Schwarz (1998), Santelli, Kaiser, Hirsch, Radosh, Simkin, and Middlestadt (2004), and Sieving, Eisenberg, Pettingell, and Skay (2006)). In addition, earlier work indicates that peer norms regarding sexual behavior are established within gender (National Research Council Panel (1987)), which supports the use of same-gender classmates as the reference group for this effect.[11] The age term (αa_{it}) is intended to capture the individual maturation process, which is both biological and psychological. The permanent individual component (ω_i) reflects aspects of the potential costs and benefits of sexual activity that vary across individuals. For example, this would capture differences in the desire for sex, as well as differences in the costs of pregnancy and sexually transmitted diseases (STDs) or the perceptions of these risks.

In the model, individuals are forward looking and consider future payoffs with a discount rate β. This is consistent with strong evidence of anticipation and intentionality in sexual initiation that was found by Kinsman et al. (1998). Consequently, because the model ends with high school graduation but the payoff to virginity status continues, nontrivial terminal values are needed. For nonvirgins, I eliminate the peer influence on preferences after high school (there is no further data, anyway), and hold the age and

---

[9]It might be preferable to estimate a more flexible matching function that does not require either of these assumptions about the number of matches per period. However, this would ask more from the available variation in the data, and the additional parameters would increase the computational burden.

[10]To see this, note that the per-period utility for an arbitrary individual (virgin or nonvirgin) can be expressed as $y_{it} \cdot u(a_{it}, Y_{i,t-1}, \omega_i, \varepsilon_{it}) + (1 - y_{it}) \cdot 0$. This shows the interaction between the individual outcome (y_{it}) and the peer-group outcome ($Y_{i,t-1}$) in the preferences.

[11]Also, the use of *lagged* peer nonvirginity rates is supported by work such as Kinsman et al. (1998) that focuses on perceptions about how many peers are *already* sexually experienced. However, as explained in Section 4.2.2, this timing is not crucial for the identification of the model.

permanent individual components constant for an infinite horizon. This yields a simple terminal value of $(\alpha A + \omega_i)/(1 - \beta)$. For virgins, the terminal value is a free parameter $\nu(\omega_i)$. This is nonzero to allow virgins to anticipate a payoff from sexual activity later in life.[12]

Lifetime discounted payoffs are thus

$$\sum_{t=1}^{T} \beta^{t-1} y_{it} u(a_{it}, Y_{i,t-1}, \omega_i, \varepsilon_{it}) + \beta^T \left(y_{iT} \frac{\alpha A + \omega_i}{1 - \beta} + (1 - y_{iT}) \nu(\omega_i) \right).$$

This can be expressed recursively using the Bellman representation, with age-specific value functions denoted $V_a(y_{t-1}, Y_{t-1}, \omega, \varepsilon)$. The vector Y_{t-1} (8×1) contains the non-virginity rates by gender in each of the four grades in high school; this is the aggregate state of the local market. For a nonvirgin, who is in the absorbing state, the value function has an analytical expression as

$$V_a(1, Y_{t-1}, \omega_i, \varepsilon_{it}) = \hat{u}_{it} + \omega_i + \varepsilon_{it} + \sum_{s=1}^{A-a} \beta^s [E_t \hat{u}_{i,t+s} + \omega_i]$$

$$+ \beta^{(A-a+1)} \frac{\alpha A + \omega_i}{1 - \beta}, \tag{3}$$

where $\hat{u}_{it}$ is the common component of payoffs defined in (2) and E_t denotes the individual's expectation given the information set at time t. For a virgin, the value function is a more complicated object that incorporates the search decision and the arrival rate. It is expressed as

$$V_a(0, Y_{t-1}, \omega_i, \varepsilon_{it})$$

$$= \max_{d_{it}} d_{it} E_t \big[\lambda_{it} \cdot \big(\hat{u}_{it} + \omega_i + \varepsilon_{it} + \beta V_{a+1}(1, Y_t, \omega_i, \varepsilon_{i,t+1}) \big)$$

$$+ (1 - \lambda_{it}) \cdot \beta V_{a+1}(0, Y_t, \omega_i, \varepsilon_{i,t+1}) \big] \tag{4}$$

$$+ (1 - d_{it}) \beta E_t V_{a+1}(0, Y_t, \omega_i, \varepsilon_{i,t+1}).$$

The first two lines on the right-hand side in (4) expresses that an individual who searches ($d_{it} = 1$) will become a nonvirgin with probability λ_{it} and will remain a virgin with probability $(1 - \lambda_{it})$. The last line gives the value of not searching, in which case the individual advances to the next period still a virgin.

To form the expected values in (3) and (4), individuals need beliefs over the sequences of nonvirginity rates among peers ($Y_{it}, Y_{i,t+1}, \ldots$) and arrival rates ($\lambda_{it}, \lambda_{i,t+1}, \ldots$). In fact, beliefs over the evolution of the vector Y_t (the nonvirginity rates by gender and grade) are sufficient for both. This is because arrival rates can be derived from the decision rule for the opposite gender. The search decisions among individuals of the opposite gender (say, individual j) depend on their state variables ($a_{jt}, y_{j,t-1}, Y_{t-1}, \omega_j, \varepsilon_{jt}$). Given Y_{t-1}, it is possible to integrate the decision rule over the distributions of ω_j and

---

[12]Because only differences in payoffs are identified by choice behavior, the estimated $\nu(\omega)$ may capture omitted aspects of the terminal values for nonvirgins such as expectations about future peer norms.

ε_{jt}, along with the individual virginity statuses $y_{j,t-1}$ that correspond to the group non-virginity rates in Y_{t-1}. This yields a distribution of N_{it}, the share of searchers among the opposite gender, which in turn gives the distribution of λ_{it} based on (1).[13]

For the beliefs about the evolution of Y_t, I use an approximation to fully rational beliefs that is similar to the approaches in Krusell and Smith (1998) and Lee and Wolpin (2006). In the approximation, the distribution of Y_t given past values is Markovian, and its expected value is autoregressive with the specification

$$\mathrm{E}[Y_{kt}|Y_{t-1}] = \psi_{0k} + \psi_1 Y_{k,t-1} + \psi_2 Y_{k,t-1}^2 + \sum_{j \in s(k)} \psi_{3j} Y_{j,t-1}. \tag{5}$$

Here k indicates one element of the vector (i.e., one gender–grade group), and $s(k)$ collects the subscripts for the opposite-gender groups, which I refer to as supply groups. The nonlinear vector autoregression that stacks these elements is denoted $\psi(Y_{t-1})$. As in Krusell and Smith (1998) and Lee and Wolpin (2006), this approximation fits the true evolution of the aggregate state extremely well (see Section 5). There are two details in the implementation of these beliefs. First, because school populations are finite in the model, the approximation incorporates the impact of an individual's choice and outcome on his or her own group's nonvirginity rate.[14] Second, because the aggregate state does not contain information on cohorts not yet in high school, the nonvirginity rates for each new cohort of ninth graders are predicted based on the previous cohort.[15]

Finally, the expected costs and benefits of sexual activity embodied in the permanent individual preference term ω may relate to the probability of initiation prior to the ninth grade. Because these baseline nonvirginity rates vary across schools, the model must account for initial conditions. To do this, I specify a distribution of ω for *virgins at the beginning of ninth grade* that is conditional on the vector Y_0, which includes the nonvirginity rates among rising ninth graders just before they enter high school. There are two reasons to think that the distribution of ω among virgins might not be independent of the initial nonvirginity rates in Y_0. First, if ω is correlated among peers, then a high Y_{i0} (the proportion of nonvirgins in the gender–grade peer group) indicates a higher ω_i for the individual. Second, if ω is uncorrelated but there are common opportunities to initiate sexual activity prior to the ninth grade, the distribution of ω among the remaining virgins is affected by selection.

This distribution of ω (i.e., among virgins at the beginning of ninth grade, conditional on Y_0) is specified with a multinomial logit where $\omega \in \{\omega^k\}_{k=1}^{\kappa}$. This embodies an assumption that there are κ "types" of individuals when it comes to sexual initiation. The simplest version of the specification uses no other variables. To add variation from

---

[13]How I implement this is explained in Section 2.2 and Appendix A.1.

[14]There is a straightforward modification to (5) to account for a known value of y_{it} in Y_{it}, in a group of given size n_i. The approximation ignores any impact on other groups.

[15]I use the nonvirginity rate of one cohort in the summer after ninth grade (e.g., $t = 4$ for the reference cohort) to predict the rate for the new cohort in the same time period. I do this by inverting the following regression for the annual growth of nonvirginity rates during ninth grade: $\mathrm{E}Y_{k4} = \Pi_0 + \Pi_1 Y_{k0}$ (k denotes a gender-cohort group). The formula for the prediction is then $\widehat{Y}_{k'4} = Y_{k4}/\Pi_1 - \Pi_0/\Pi_1$, where k' denotes the new ninth-grade cohort.

exogenous characteristics that may relate to the expected costs and benefits of sex, and thereby improve the precision of the estimates, a second version includes a vector of permanent individual-level observables, x. A third version further adds the means of these observables within the local market, $\bar{x}$. This turns out to be similar to including school fixed effects, in terms of the identification of the model (see Section 4.2 for details).[16] These three alternative versions of the conditional distribution of ω can be expressed as

$$\Pr(\omega = \omega^k | Y_0, z) = \pi_{k|Y_0, z} \equiv \frac{\exp(\pi_0^k + Y_0' \pi_1^k + z' \pi_2^k)}{1 + \sum\limits_{l=2}^{\kappa} \exp(\pi_0^l + Y_0' \pi_1^l + z' \pi_2^l)}, \tag{6}$$

where z is either empty, x, or $(x, \bar{x})$, respectively, in the first, second, and third versions. To interpret ω, it is important to keep in mind that while this variable is realized at the individual level, it is intended to capture the effects of both individual- and group-level factors on preferences about sex. The values of ω are naturally correlated among the students in a school due to the presence of Y_0 (and $\bar{x}$, in the third version) in the expression above.

2.2 Solving the model

Given beliefs about the evolution of Y_t, the individual decision problem solves much like a standard single-agent dynamic problem. The expression for the value function in (4) can be rearranged to

$$\begin{aligned} &V_a(0, Y_{t-1}, \omega_i, \varepsilon_{it}) \\ &= \max_{d_{it}} d_{it} \cdot \mathrm{E}_t \lambda_{it} \cdot \big(\hat{u}_{it} + \omega_i + \varepsilon_{it} + \beta \mathrm{E}_t V_{a+1}(1, Y_t, \omega_i, \varepsilon_{i,t+1}) \\ &\quad - \beta \mathrm{E}_t V_{a+1}(0, Y_t, \omega_i, \varepsilon_{i,t+1}) \big) + \beta \mathrm{E}_t V_{a+1}(0, Y_t, \omega_i, \varepsilon_{i,t+1}). \end{aligned} \tag{7}$$

Because $\mathrm{E}_t \lambda_{it}$ is strictly positive, the decision rule is, therefore,

$$\begin{aligned} d_{it} = 1 \quad \text{iff} \quad &\hat{u}_{it} + \omega_i + \varepsilon_{it} + \beta \mathrm{E}_t V_{a+1}(1, Y_t, \omega_i, \varepsilon_{i,t+1}) \\ &> \beta \mathrm{E}_t V_{a+1}(0, Y_t, \omega_i, \varepsilon_{i,t+1}). \end{aligned} \tag{8}$$

Thus a virgin will search if and only if the value of becoming sexually active exceeds the value of remaining a virgin. This is a standard result in a model with no search cost.

It is important to note that the availability of partners still has a direct impact on the search decision. This is because beliefs about partner availability in the future influence the value of remaining a virgin, that is, the $V_{a+1}(0, \ldots)$ on the right-hand side of (8). As one might expect, the option value of entering the next period as a virgin is weakly

---

[16]By contrast, including school fixed effects directly would be problematic because they would drastically increase the number of parameters and could raise an incidental parameters problem.

increasing in the probability of finding a partner at that time.[17] Hence, all else equal, if the probability of finding a partner in the next period is lower, the incentive to search in the current period is greater. This has the implication that an individual could decide to search for a sexual partner even if his or her current flow value of nonvirginity ($\hat{u}_{it} + \omega_i + \varepsilon_{it}$) is negative.

The age-specific value functions for virgins, given by (7), do not have analytical expressions, but they can be numerically constructed by backward recursion. I use interpolation to approximate these functions (Keane and Wolpin (1994)) because the state space includes an 8-dimensional continuous vector (Y_{t-1}). This involves evaluating the functions on a set of points in the state space and then regressing these values on transformations of the state variables to create very close approximations to the true functions. To choose solution points that span the state space, I draw Y_{t-1} from a joint uniform distribution and ω from the set of values $\{\omega^k\}$, and sample x and the membership of the peer and supply groups from their joint empirical distribution.

To evaluate expression (7) at the solution points, I need to extend the standard procedure so as to account for the search decisions of opposite-gender virgins that are embedded in the arrival rate (λ_{it}). An exact calculation for the expected arrival rate ($E_t\lambda_{it}$) would use the decision rule in (8), and integrate over the values of ω and ε_t for opposite-gender virgins. However, the random values of Y_{t-1} drawn for the solution points do not correspond to the individual virginity statuses of the members of the supply groups, and there is no simple procedure to choose virgins and nonvirgins to match Y_{t-1}. This is because the probability of being a nonvirgin at period $t-1$ depends on ω and the entire history of Y. Instead, I approximate the search decisions among the opposite gender as a function of Y_{t-1} and use this to approximate the expected arrival rate. This procedure is described further in the Appendix A.1.

Equilibrium beliefs about the evolution of Y_t are recovered directly from the data. This follows methods introduced by Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), and Pakes, Ostrovsky, and Berry (2007). In my case, the autoregression ψ in (5) is estimated in a preliminary stage (but not the individual choice probabilities, because I use backward recursion). As in the above papers, this approach assumes only one equilibrium is observed, and it assumes a steady state from one cohort to the next. Moreover, because I use an approximation to rational beliefs, unlike these papers, I need to check that the estimated beliefs are consistent with the model. I do this by reestimating ψ on data simulated from the model post-estimation and comparing the two estimates of ψ with each other. The results support the approximation (see Section 5). Also, because the autoregression fits the observed evolution of Y_t extremely well, with $R^2 > 0.95$, I use a degenerate distribution at the expected values for the beliefs in the approximation.[18] This avoids the need to integrate over a distribution in each future pe-

---

[17]To see this, use (7) to expand the expression for $V_{a+1}(0,\ldots)$. If the value inside the parentheses (the part that corresponds to the decision rule in (8)) is positive, then $d_{i,t+1} = 1$ and so $V_{a+1}(0,\ldots)$ is increasing in $E_t\lambda_{i,t+1}$, the expected arrival rate in the next period. If the value inside the parentheses is negative, then $d_{i,t+1} = 0$ and so $V_{a+1}(0,\ldots)$ is invariant to $E_t\lambda_{i,t+1}$.

[18]Krusell and Smith (1998) and Lee and Wolpin (2006) assume large numbers of agents, so that the evolution of the aggregate state is deterministic in their cases.

riod when solving for individual behavior. There is a bias, of course, because the value functions for virgins are nonlinear, but it should be small given the tightness of the distribution around the predicted values.

An alternative to the two-step estimation procedure would be to solve for the approximation ψ as a fixed point along with the structural parameters, as in Lee and Wolpin (2006). In that paper, part of the aggregate state is unobserved to the econometrician (there is an aggregate productivity shock), so it is not possible to estimate an approximation to rational beliefs directly from the data. Given that the aggregate state for my model *is* observed, the advantage of recovering beliefs directly from the data is that it avoids the iteration needed to solve a fixed point for each candidate set of structural parameters. This greatly reduces the computational burden of estimation.

3. Data and descriptive statistics

The data come from Waves I and II of the National Longitudinal Study of Adolescent Health (Add Health). The study contains a nationally representative sample of students in grades 7–12 during the 1994–1995 school year, when the first wave was conducted. The second round of interviews (Wave II) followed up with respondents 1 year later in April–August 1996. Add Health features a highly clustered sample drawn from 80 high schools plus additional middle schools that feed students into the sample high schools (one middle school per high school, unless the sample high school already includes grades seven and eight).

Add Health collects detailed retrospective histories on sexual activity and romantic relationships. To enhance the sense of privacy, these questions were administered in a self-directed portion of the survey on a laptop computer at respondents' homes. Included in these questions, respondents are asked if they have ever had sexual intercourse, which is defined explicitly.[19] Those who say yes are then asked to report the month and year of first sex. Both rounds of interviews ask these questions of all respondents, and to minimize the loss of observations due to missing data, I use the earliest date reported in either round. From these observations, I construct a quarterly series on virginity status for each individual, starting in the summer of 1994 and ending in the spring of 1996.

The estimation sample uses individuals observed in grades 9–12 in either the 1994–1995 or 1995–1996 school years, who were selected for in-home interviews.[20] Add Health contains 17,657 such individuals, who are in grades 8–12 during the first round of interviews in 1994–1995. I use the grade in that academic year to refer to separate "cohorts." I exclude 2635 individuals who drop out of the second round of interviews (except for the twelfth grade cohort, which was not reinterviewed). I also exclude 69 individuals

---

[19]The question reads, "Have you ever had sexual intercourse? When we say sexual intercourse, we mean when a male inserts his penis into a female's vagina." (Wave I Adolescent In-Home Questionnaire Code Book, Section 24, p. 1.)

[20]Add Health also administered an in-school questionnaire to all students in the sampled schools.

(a) Nonvirginity rates, boys

(b) Nonvirginity rates, girls
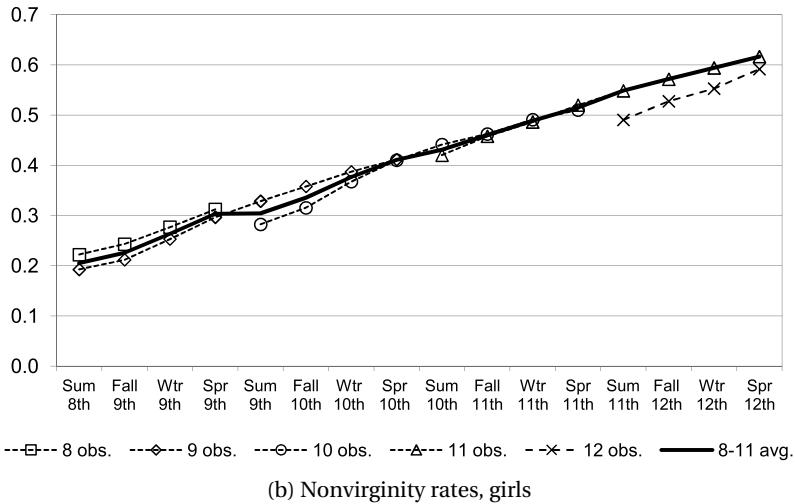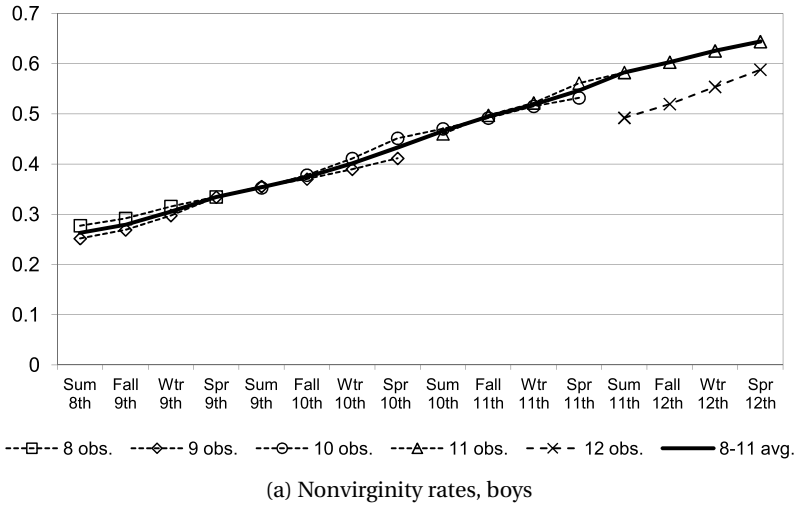
FIGURE 1. Observed nonvirginity rates by quarter within grade in high school. Each line "*x* obs." shows the nonvirginity rate of an individual cohort, defined by grade in the 1994–1995 school year, and "8–11 avg." averages across cohorts to make a synthetic cohort.

from an all-boys school, 98 in schools with small samples that do not have both genders in some grades, and 318 who report homosexual sex. After dropping observations without information on key identifying variables (school, grade cohort, and gender), the final estimation sample contains 14,294 individuals in 78 schools. With five cohorts per school, this means there is an average of 18 sampled individuals (median of 15) in each peer group as defined by gender and grade.

Figure 1 presents the nonvirginity rates for this sample by quarter in high school (i.e., "age" in my model). Each cohort, which is observed for 1 or 2 years, is shown as a separate line positioned over the appropriate ages. The black line averages among all individuals at each age to produce a complete path through high school for a synthetic

TABLE 1. Sample shares with given characteristics.

Variable	Unweighted Share	Weighted Share
Black	0.217	0.162
Younger child	0.500	0.489
Only child	0.190	0.198
Parent with 16+ years of edu.	0.278	0.270

cohort.[21] These graphs show that a large portion of individuals initiate sex during high school. The share of nonvirgins among boys increases from just over 26% at the beginning of ninth grade to just under 64% at the end of twelfth grade, and among girls it increases from 20% to 62%. Thus, about 40% of the population initiates sex during the four years of high school.

Data on the characteristics (x) that appear in the distribution of the individual preference term (ω) come from Wave I. I use black race, parental education, and sibling status for this, because they are predetermined and have been shown to predict age of sexual initiation in prior work.[22] The education variable indicates whether one parent has 16 or more years of education, and the sibling variables are two dummies for being a younger sibling and being an only child. Table 1 gives the unweighted and weighted means of these indicators (i.e., the sample shares). The weights make little difference except for the share with black race, which reflects oversamples in the sample design.

Table 2 shows the raw correlation between individual virginity status and the non-virginity rates for each gender and grade at the same school, assessed in the last observation period (the spring quarter of 1996).[23] The boldface numbers along the diagonal give the correlations with the same-gender–same-grade peer groups. These are somewhat higher than the correlations with other grades of the same gender (except for girls in the tenth and twelfth grades, who have slightly higher correlations with some other grade), which provides support for the definition of peer groups by grade (in addition to gender). More broadly, there are large correlations in virginity status within schools, about 0.2 in magnitude, which indicates the substantial variation in nonvirginity rates across schools.

The boldface elements of the cross-gender blocks in Table 2 indicate the "supply groups." These are the grades used for the endogenous supply of partners in the empirical implementation. For boys, the supply groups are girls in the same grade, the grade below, and the grade above; for girls, they are boys in the same grade and the next two

---

[21]I exclude the twelfth grade cohort from the synthetic cohort because they are interviewed only once, so they have a higher rate of missing data on the month of first sex. This makes their retrospective nonvirginity rates fall below the trend constructed from the younger cohorts.

[22]See, for example, NRC Panel (1987) and Levine (2001) on various individual and family characteristics that predict early sexual initiation, and Widmer (1997) and Argys, Rees, Averett, and Witoonchart (2006) on the influence of siblings.

[23]The individual is excluded from the nonvirginity rate for his or her own peer group.

TABLE 2. Correlation between individual virginity status and nonvirginity rates of each gender–grade group at the same school.

| Individual | Comparison Group | | | | | | | |
| | Boys | | | | Girls | | | |
Gender and Grade	Grade 9	Grade 10	Grade 11	Grade 12	Grade 9	Grade 10	Grade 11	Grade 12
Boys								
9	**0.267**	0.229	0.236	0.169	**0.243**	**0.286**	0.190	0.085
10	0.161	**0.216**	0.204	0.116	**0.144**	**0.224**	**0.154**	0.110
11	0.190	0.189	**0.198**	0.149	0.157	**0.188**	**0.151**	**0.119**
12	0.127	0.107	0.144	**0.157**	0.123	0.144	**0.142**	**0.190**
Girls								
9	**0.212**	**0.207**	**0.208**	0.165	**0.188**	0.173	0.156	0.137
10	0.216	**0.214**	**0.200**	**0.144**	0.114	**0.193**	0.199	0.156
11	0.154	0.155	**0.157**	**0.155**	0.094	0.210	**0.220**	0.176
12	0.052	0.116	0.113	**0.191**	0.105	0.150	0.176	**0.168**

Note: The individual is excluded from the nonvirginity rate for his/her own group. Peer and supply groups shown in bold.

older grades.[24] These were chosen because, in the sexual histories, more partners are reported from these grades than any others. The purpose of these restrictions is to incorporate the low probability of matches between certain grades without adding further complexity to the model. Partners from outside these grades, such as an eleventh grade girl for a ninth grade boy or vice versa, are considered to be exogenous, which treats them as part of the external market. With a few exceptions, the correlations in virginity status between individuals and their designated supply groups are larger than the correlations with the excluded grades.[25]

3.1 *Evidence of the composite effect*

Before turning to the estimation of the search and matching model, I briefly present evidence of a composite effect of social interactions in the framework of a simple haz-

[24]The supply groups do not need to be symmetric because the lack of constraint on the number of partners makes it possible for a small number of individuals from one grade to match with a large number from another grade.

[25]Clearly the correlations with some of the excluded grades of the opposite gender are very close to the correlations with the included grades. The same could be said about the correlations with other grades of the same gender, which are similarly excluded from the peer group. Peer and supply effects from these groups would likely be smaller than from the included groups, but they may be nontrivial. A more flexible approach would be to have separate parameters for the effects of peer norms and partner availability for each pairwise combination of grades. Then γ and λ_1 would have 16 elements each for boys and girls, for a total of 64 parameters, which would greatly increase the computational burden. Here instead λ_1 is made partially flexible, with one parameter for each of the three supply groups, so as to facilitate the counterfactual that isolates the ninth grade from the rest of high school.

ard model for sexual initiation.[26] In addition to demonstrating the presence of social effects, this exercise illustrates how the initial nonvirginity rates and other variables in the distribution of preference types ((6)) capture heterogeneity that otherwise would be attributed to endogenous social effects. Also it is possible here to include further variables that provide a partial assessment of the identifying assumptions discussed below in Section 4.2.2.[27] Specifically, I include information on school policies related to family planning, because these policies might not be captured with the initial nonvirginity rates if they are shaped by factors within the high schools themselves, independently of any influences on sexual activity prior to high school. I also estimate models with school fixed effects to see how the results compare in this finite sample when either fixed effects or initial nonvirginity rates are used to control for common unobservables within schools.

The models are specified with discrete-time logit hazard functions.[28] They include age, the lagged nonvirginity rates in the reference group(s) (either the peer group alone or the peer and supply groups), the initial nonvirginity rates, and additional variables that are added in sequence.[29] Those variables are the individual characteristics x (as in the second version of (6)), the group means of these characteristics $\bar{x}$ (as in the third version of (6)), and indicators for seven school policies on sex education and family planning.[30] To be clear, these hazard models cannot be interpreted as approximations of the search and matching model, as there is no direct relationship between their coefficients and the parameters of my model. Still the exercise is informative because the coefficients on the lagged nonvirginity rates capture essentially the same variation that identifies the social effects in my model: changes in nonvirginity rates over time conditional on their initial rates.

The estimates of the composite social effects from these models are presented in Table 3. These are the coefficients on the lagged nonvirginity rates in the reference group(s). (The full results from each specification are in Tables A.1–A.3 in the Appendix.) There are three panels for three different definitions of the reference group(s): first the same-gender–same-grade peer group alone, then the combination of the peer and supply groups, and finally the peer and supply groups treated separately. Columns 1 (boys) and 7 (girls) first report estimates from models with only age and the lagged nonvirginity rates in the reference group(s). When the initial peer and supply group nonvirginty rates (Y_{i0} and $Y_{s(i),0}$) are added in columns 2 and 8, the coefficients on the lagged nonvirginity rates are drastically reduced. They fall by half or more in magnitude, and for girls, the coefficient on the supply group nonvirginity rate is essentially eliminated in

---

[26]Additional evidence of a composite effect, using a static regression framework with very different identifying assumptions, can be found in an earlier version, *PIER Working Paper* 12-015, available at http://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/12-015.pdf.

[27]I am grateful to a referee for suggesting this aspect of the analysis.

[28]In other words, $\Pr(y_{it} = 1 | y_{i,t-1} = 0, z_{it}) = \Lambda(\theta' z_{it})$, where z_{it} are the explanatory variables.

[29]For this exercise, I use only one supply group, which is the opposite gender in the same grade.

[30]The underlying policies are the grade(s) when sex education is offered; whether counseling on family planning is available at the school, is offered via referral to an outside provider, or is neither provided nor referred by the school; whether daycare is provided for the children of students; and whether for-credit courses in parenting are offered to pregnant students.

TABLE 3. Composite social effects in logit hazard models.

Variable	Boys						Girls					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Same-Gender Peer Group												
Nonvirginity rates												
Peer group lag	1.37	0.59	0.53	0.62	0.54	0.42	1.25	0.57	0.51	0.49	0.29	−0.36
$(Y_{i,t-1})$	(0.22)	(0.42)	(0.42)	(0.42)	(0.43)	(0.32)	(0.19)	(0.32)	(0.32)	(0.33)	(0.33)	(0.29)
Initial rates incl.	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	no
$(Y_{i,0}$ and $Y_{s(i),0})$												
Panel B: Peer and Supply Groups Combined												
Nonvirginity rates												
Peer and supply lag	1.74	0.74	0.69	0.79	0.70	0.69	1.51	0.64	0.60	0.60	0.41	−0.18
$((Y_{i,t-1}+Y_{s(i),t-1})/2)$	(0.24)	(0.44)	(0.44)	(0.45)	(0.45)	(0.40)	(0.21)	(0.39)	(0.40)	(0.41)	(0.41)	(0.36)
Initial rates incl.	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	no
$(Y_{i,0}$ and $Y_{s(i),0})$												
Panel C: Peer and Supply Groups Separately												
Nonvirginity rates												
Peer group lag	1.00	0.42	0.36	0.45	0.38	0.35	0.95	0.55	0.47	0.43	0.22	−0.40
$(Y_{i,t-1})$	(0.24)	(0.46)	(0.46)	(0.46)	(0.46)	(0.32)	(0.22)	(0.35)	(0.35)	(0.36)	(0.36)	(0.29)
Supply group lag	0.75	0.34	0.33	0.35	0.33	0.34	0.58	0.05	0.11	0.15	0.19	0.19
$(Y_{s(i),t-1})$	(0.23)	(0.37)	(0.37)	(0.37)	(0.37)	(0.30)	(0.20)	(0.39)	(0.39)	(0.39)	(0.39)	(0.28)
Initial rates incl.	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	no
$(Y_{i,0}$ and $Y_{s(i),0})$												
All Panels												
Indiv. chars. incl.	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes
Peer means incl.	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
School policies incl.	no	no	no	no	yes	no	no	no	no	no	yes	no
School fixed effects	no	no	no	no	no	yes	no	no	no	no	no	yes
Observations	21,671	21,671	21,671	21,671	21,671	21,504	24,116	24,116	24,116	24,116	24,116	24,045

Note: Standard errors are given in parentheses.

panel C. This indicates that the initial nonvirginity rates do indeed capture a substantial amount of variation from common unobservables that would otherwise bias the estimates of social effects. Columns 3 and 9 add the individual characteristics, and columns 4 and 10 add the group means of these characteristics. These make little difference for the estimated coefficients on the lagged nonvirginity rates. In particular, this suggests that student selection into high schools (as captured with the group means) does not bias the estimates of social effects if the initial nonvirginity rates are included. Columns 5 and 11 add the school policy indicators. For boys, the coefficients on the lagged nonvirginity rates change very little. For girls, the coefficient on the peer group rate is reduced by one-half in panel C. This change appears to be driven by two policy indicators that have strong and possibly counterintuitive associations with the initiation hazard (shown in Tables A.1–A.3): having sex education in grades 11 or 12 (a positive association) and offering family planning services at school (a negative association).[31] Finally, columns 6 and 12 drop the initial nonvirginity rates and, instead, use school fixed effects to control for common unobservables. The estimates of the composite social effects for boys show only small changes, but the estimates for girls are reversed in sign. However, there appears to be a substantial problem with overfitting the data when school fixed effects are used in this finite sample. The coefficient on age increases drastically (Tables A.1–A.3), which indicates that the fixed effects can be set to predict very closely when the observed exits occur in each school.[32]

Overall these results provide strong evidence that school-based social interactions have an effect on the timing of sexual initiation for both boys and girls. A coefficient of about 0.5 on the lagged nonvirginity rates, for example, implies an average marginal effect of 0.022 (0.024) for boys (girls), so a 1-standard-deviation increase in the rates would raise the initiation hazard by 8.7 (8.5)% in relative terms.[33] Moreover, this exercise broadly supports the identification strategy. The coefficients on the lagged peer and supply group nonvirginity rates are fairly stable once the initial rates are included, which suggests that any biases from factors not captured with the initial rates would not qualitatively affect the results from the search and matching model.

4. Structural estimation

I estimate the search and matching model via maximum simulated likelihood. The model generates a discrete-time duration to first sexual intercourse, based on the prob-

---

[31]Oettinger (1999) shows theoretically how sex education could either increase or decrease the propensity to initiate sex by changing expected payoffs and risk probabilities, and he presents evidence that sex education increases the initiation hazard among girls. However, in the estimates shown here it is hard to reconcile the positive association with initiation from sex ed in grades 11 and 12, and the lack of association from sex ed in grades 9 and 10 (Tables A.1–A.3).

[32]Note that individual fixed effects could not be included in a discrete-time logit hazard model with a strictly monotonic variable such as age. Exits necessarily occur in the last period an individual is observed, so the fixed effect for each individual would be driven toward negative infinity, while the coefficient on age would be driven toward positive infinity.

[33]These figures are calculated using the models in columns 3 and 9 of panel A, which have coefficients on the lagged nonvirginity rates that are close to 0.5. The standard deviations of the peer group nonvirginity rates are 0.18 and 0.19 for boys and girls, respectively, and the average hazard rates are 0.046 and 0.053.

ability each period of the transition from virginity to nonvirginity. This transition probability is the product of the probabilities of searching and of finding a partner; however, search decisions are unobserved. To separate the search probability from the arrival rate without relying on functional form, I take advantage of additional data on the arrival of subsequent partners after the first. Accordingly, the likelihood function includes individual contributions for both the duration to first sex and the arrival of subsequent partners.

4.1 *Likelihood function*

The likelihood contributions for the durations to first sex take the form of a finite mixture because the permanent component of preferences, ω, has a discrete distribution. Conditional on ω, the per-period transition probability is the product of the arrival rate and the probability that the decision rule in (8) is satisfied.[34] With ε distributed standard normal and its CDF denoted Φ, this product is

$$L_{it}(\omega) \equiv \Phi\big[\hat{u}_{it} + \omega + \beta E_t V_{a_{it}+1}(1, Y_t, \omega, \varepsilon_{i,t+1}) - \beta E_t V_{a_{it}+1}(0, Y_t, \omega, \varepsilon_{i,t+1})\big]$$
$$\cdot \int \lambda_{a_{it}}(N_{it}) f(N_{it}) \, dN_{it}. \tag{9}$$

The solution to the model for a particular set of parameters provides the expected future values of virginity and nonvirginity inside Φ. Simulation is needed for the arrival rate (the second line above), so as to integrate over the unobserved search decisions of opposite-gender virgins, which generate N_{it}. The simulation procedure is described in Appendix A.2.

For an individual who initiates sex in period t_i^*, the type-specific probability for the observed duration is[35]

$$L_i(\omega) \equiv L_{it_i^*}(\omega) \cdot \prod_{s=1}^{t_i^*-1} \big[1 - L_{is}(\omega)\big]. \tag{10}$$

Adding across types, the probability given for the duration is

$$\mathcal{L}_i \equiv \sum_{k=1}^{\kappa} \pi_{k|Y_{m0}, z_i} L_i(\omega^k), \tag{11}$$

where $\pi_{k|Y_{m0}, z_i}$ is defined in (6) and m is the individual's school (i.e., market).

In addition to the above durations to first sex, the likelihood function contains individual contributions from nonvirgins for the arrival of their subsequent partners, so as to identify the arrival rate parameters λ_0 and λ_1. This employs data from the detailed sexual histories that report when sex first occurred with each partner. The use of nonvirgins exploits the fact that, in the model, they search every period, so the arrival of

---

[34]The transition probability factors in this way because the remaining unobservable in the search decision is the IID preference shock ε_{it}.

[35]For individuals who are not observed to initiate sex, this is $\prod_{s=1}^{T}[1 - L_{is}(\omega)]$.

subsequent sexual partners after the first one directly identifies the raw arrival rate. To limit departures from the model, specifically the assumption that partner arrival rates are the same for virgins and nonvirgins, I only use the arrival of second partners for this purpose. The estimated arrival parameters will be biased to the extent that arrival rates of second partners differ from arrival rates of first partners, and this bias could go in either direction. Exclusivity in relationships would reduce the arrival rate of second partners because individuals do not immediately continue to search once they have a first partner. On the other hand, learning how to meet partners would increase the arrival rate. Any bias is partially mitigated, however, because the arrival rate function also appears in the likelihood contributions for the durations to first sex.

The individual likelihood contribution for the arrival of a second partner is

$$
\mathcal{A}_i \equiv \mathrm{E}\big[\lambda_{a_{it_i^{**}}}(N_{it_i^{**}})\big] \prod_{s=t_i^*}^{t_i^{**}-1} \big(1 - \mathrm{E}\big[\lambda_{a_{is}}(N_{is})\big]\big), \tag{12}
$$

where t_i^{**} is the period when sex first occurred with the second partner and $\mathrm{E}[\lambda(N)]$ is the integral over the distribution of N, as in the second line of (9).[36] I restrict to individuals with $y_{i0} = 0$ (initial virgins) so as to observe the beginning of these spells.

Finally, because the arrival of each partner is assumed to be an independent event and to be independent of individual characteristics, the likelihood contributions in (12) simply multiply with the likelihood contributions in (11). Thus the complete log-likelihood function is $\sum_i \log(\mathcal{L}_i) + \sum_i \log(\mathcal{A}_i)$, using individuals who are virgins at $t = 0$.

The estimation sample includes individuals in cohorts that are first observed after the ninth grade ($a_{i0} > 0$). This presents a dynamic selection problem because individuals who are still virgins in later grades are more likely to have low values of ω. The estimation procedure needs to account for this; however, because the hazard rate is a function of time-varying arguments, there is not a simple way to integrate over the unobserved periods. Instead, to update the distribution of ω for virgins in cohorts that are first observed after the ninth grade, I use data from younger cohorts at the same school to create approximate, type-specific hazard rates. With these I can calculate the probability, for each type, of still being a virgin when the individuals are first observed, and then update the initial distribution of ω (for ninth graders) via Bayes rule. The exact procedure is described in Appendix A.3.

For the standard errors, I use the asymptotic distribution of a standard maximum likelihood estimator. This assumes that the number of simulations for the expected arrival rates grows fast enough with the sample size (Gourieroux and Monfort (1997)). The variance approximation is calculated via numerical differentiation.[37]

---

[36]The first period when sex occurred with the first partner is included in the duration to the second partner because multiple partners are possible per period.

[37]The variance estimate ignores the first-stage estimation of the beliefs approximation. Murphy and Topel (2002) discuss this issue and propose an estimator, but it would be cumbersome to implement with a dynamic, structural model.

4.2 *Identification*

The analysis of identification has two parts. First I provide a formal argument for the identification of the structural parameters. Then I discuss more generally how my empirical approach addresses identification problems that typically arise in social interactions models.

Before the formal discussion, however, it is worth noting why the estimation of a dynamic model with longitudinal data would be required to separately identify the effects of peer norms and partner availability. If we believe these two mechanisms interact over time, as in my model, then it would be impossible to distinguish between them in a static model with cross-sectional data. Consider the effect of an exogenous increase in the nonvirginity rate among same-gender peers. The change in peer norms would directly increase an individual's demand for sex. However, the increase in the peer non-virginity rate would also indirectly increase the supply of partners over time. This indirect effect is easily seen in a dynamic context: in period 1, the nonvirginity rate among peers exogenously increases; the search behavior of these additional nonvirgins raises the arrival rate of partners for the opposite gender; this results in a higher nonvirginity rate among the opposite gender in period 2; hence, in period 3, both the individual's search probability (demand) and arrival rate (supply) are higher. In a static framework, these effects cannot be disentangled. Put in more general terms, the problem is that any exogenous shift in demand indirectly shifts the supply curve over time as well. Hence, exogenous changes in demand cannot be used to trace out the supply curve, and vice versa, at least not in a cross section.

4.2.1 *Identification of the structural parameters*   The observed data consist of permanent individual characteristics, x, virginity status over time, $(y_t)_{t=0}^{T}$, and the arrival date of second partners, t^{**}, along with "age" (grade × quarter) and gender. Individuals (i) are grouped together into schools (m), so that the sample consists of $\{(x_{mi}, (y_{mit})_{t=0}^{T}, t_i^{**}), i = 1, \ldots, n_m\}_{m=1}^{M}$, and the asymptotic argument has $M \to \infty$. In what follows, I show identification with a sequential process, although the estimation procedure has only two steps (the first step below is done separately from the others).

1. As defined earlier, let $Y_{kt} = n_k^{-1} \sum y_{kit}$, where k indexes one gender–grade group, so that $Y_t = (Y_{1t}, \ldots, Y_{8t})$ is the vector of nonvirginity rates for the eight groups in a high school. The nonlinear vector autogregression ψ in (5) is identified from the joint distribution of (Y_t, Y_{t+1}). This represents equilibrium beliefs about the evolution of Y_t under the assumptions that (a) one equilibrium is observed in the data and (b) beliefs are degenerate at their expected values.

2. Conditional transition probabilities from virginity to nonvirginity can be calculated. Let $P_a(x_i, \mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0)$ be the probability that $y_{it} = 1$ for an individual at age a, given $y_{i,t-1} = 0$ and conditional on the listed arguments. The vectors $\mathbf{x}_s = \{x_j : j \in S_a(i)\}$ and $\mathbf{y}_{s,t-1} = \{y_{j,t-1} : j \in S_a(i)\}$ contain the permanent characteristics and lagged virginity statuses of the members of the supply groups, as $S_a(i)$ collects the indices of the members of the supply groups for individual i at age a.[38]

---

[38]The variables $\mathbf{x}_s$ and $\mathbf{y}_{s,t-1}$ affect the fraction of searchers in the supply groups, N_t. Accordingly they are needed to estimate the parameters of the arrival rate function.

3. Similarly, the arrival rate of subsequent partners after the first can be calculated from the data: $Q_a(\mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0)$.[39]

4. The conditional probability that a virgin searches ($d_{it} = 1$) can then be found as $F_a(x_i, \mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0) = P_a(x_i, \mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0)/Q_a(\mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0)$. This uses the assumption that the arrival rate is the same for first sexual partners as for second sexual partners.

5. At this point, the conditional choice probabilities and all state transition probabilities are known.[40] Hence the identification of the utility parameters is the same as in a single-agent discrete choice dynamic programming model, with unobserved heterogeneity specified as a finite mixture (e.g., Keane and Wolpin (1997)). As is standard, I assume a distribution for ε, a parametric form for u, and a value for β. Because Y_0 has a continuous distribution, I also assume a parametric form for the type probabilities.

6. Finally, the parameters of the arrival rate function can be recovered. The model specifies that

$$Q_a(\mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0) = \mathrm{E}\big[\Lambda(\lambda_{0a} + \lambda_1 N_t)|\mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_0\big],$$

where Λ is the logistic function. The random variable N_t is the fraction of searchers within the supply group, $N_t = |S_a(i)|^{-1}\sum_{j \in S_a(i)} d_{jt}$, where $d_{jt} \in \{0, 1\}$ is the search decision of individual j. The distribution of each d_{jt} is now known thanks to the recovery of the structural parameters in step 5. Hence, the distribution of N_t is known, so the expectation can be computed and the equation can be solved for λ_{0a} and λ_1.[41]

At this point, all the structural parameters have been identified.

4.2.2 *The reflection, selection, and correlation problems*　　I now consider identification problems commonly raised in the social interactions literature and discuss how they are addressed beyond the use of parametric assumptions.

Manski (1993) defined the "reflection" problem, which is that linear models with social interactions are not identified if the means of both peer outcomes and peer characteristics directly affect individual behavior. In a myopic game, the use of lagged peer outcomes for the endogenous social effect would circumvent this problem, as Manski (1993) notes. However, in my model, beliefs about current and future peer outcomes also affect behavior, so the use of lagged peer nonvirginity rates in the flow utility does not simplistically resolve the problem.[42] Instead, the reflection problem does not arise for two reasons. First, the model is nonlinear. Brock and Durlauf (2001a, 2001b) show that the reflection problem does not arise in binary choice models and certain other

---

[39]The characteristics of the individual are excluded based on the model, although this does not affect the identification argument.

[40]The individual state transition is $\Pr(y_{it} = 1|d_t, \ldots, y_{i,t-1} = 0) = d_t \cdot Q_a(\mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0)$ and the aggregate state transition is $\mathrm{E}(Y_t|d_t, x_i, \mathbf{x}_s, \mathbf{y}_{s,t-1}, Y_{t-1}, Y_0, y_{i,t-1}) = \psi(Y_{t-1})$ in the approximation.

[41]The same argument applies for the alternative specification of the matching technology, which uses the ratio of opposite-gender to own-gender searchers.

[42]Brock and Durlauf (2001b) and Blume et al. (2011) provide thorough discussions about the consequences for identification from alternative assumptions about the timing of endogenous social effects.

nonlinear models because mean group outcomes are no longer linearly dependent on mean characteristics. This applies in my model as well.[43]

Second, and perhaps more robustly, the reflection problem would not arise even if there were a linear dependence because the assumption of stable preferences over time provides an important identifying restriction. In Manski's (1993) model and the related models considered in Brock and Durlauf (2001b), the exogenous group-level variables are essentially arbitrary, so there is no theory to inform the relationship between their direct effects on behavior and the effect of the expected group outcome. Separate parameters must be recovered for these exogenous and endogenous social effects, but the parameters are not separately identified when there is a linear dependence. In my model, the analogue of the exogenous social effect comes from the lagged group outcomes, while the endogenous effect comes from the expected current and future outcomes. The effects of these variables on expected utility, and hence on behavior, are the same (up to a known discount factor) because the flow utility function is the same over time. Thus there are not two separate parameters to identify. (See Appendix B.1 for a more formal discussion.)

Turning to the other challenges to identification, the definition of peer groups by grade is intended to avoid a selection problem that would occur if endogenous friendships or activities like sports teams were used.[44] However, there remains the question of whether individuals select into schools or school districts. This raises the more general problem of correlated unobservables within social groups; that is, "correlated effects" in Manski's (1993) typology. Selection into schools is one mechanism that can generate such a correlation, but there are many others. Any unobserved individual or family factors that are correlated within schools, and any unobserved common factors that arise from school attributes, school policies, or the community where the school is located, would result in correlated effects.

The approach taken here to address the problem of correlated unobservables is to have the distribution of preference types (ω) be a function of initial nonvirginity rates (Y_0). As I explain in Appendix B.2, for the purpose of identification this is equivalent to having school fixed effects in the distribution of types, under certain assumptions. In particular, this assumes that the unobserved factors are time-invariant and their effects on sexual activity are similar before high school and during high school. This would hold if the unobserved effects are primarily based on location, which I believe to be a reasonable supposition. Local economic opportunities, medical services (e.g., family planning clinics), dating markets outside the school, and local social norms (e.g., from parents rather than peers at school) could have substantial influences on the perceived

---

[43]One might be concerned that the use of an approximation to rational expectations could result in a linear dependence. However, the approximation ψ includes nonlinear terms, and Brock and Durlauf (2001b) prove that perturbations away from linear expectations facilitate identification (Theorem 7). In addition, the value function in (4) is nonlinear, so even if ψ were linear, there would not be a linear dependence between the effects of Y_{t-1} and $\psi(Y_{t-1})$ in the decision rule (8).

[44]This assumes that individuals do not systematically skip or repeat grades so as to affect their chances of sexual initiation. Hoxby (2000), Hanushek, Kain, Markman, and Rivkin (2003), and other authors similarly assume that school cohorts are exogenous with regard to their outcomes of interest.

value of nonvirginity and, hence, on decisions to become sexually active. On the other hand, this approach would not address unobserved factors that arise from high schools themselves, independently of the middle schools that feed into them or the communities where they are located. In particular, one might be concerned about the selection of students into specific high schools (rather than residential locations) and any variation in relevant school policies that is independent of location-based factors that influence sexual activity prior to high school. However, the results in Section 3.1 indicate there is little bias from factors related to the student body or school policies that are not captured by the initial nonvirginity rates. Finally, a principal advantage of my approach—that is, using initial nonvirginity rates rather than school fixed effects—is that it avoids the need to estimate a separate parameter for each school, which would drastically expand the parameter space and could raise an incidental parameters problem.

With this approach, the variation that is used to identify the effects of peer norms and partner availability comes from differences across schools in peer and supply group nonvirginity rates (Y_{it} and $Y_{s(i),t}$), conditional on their initial rates (Y_{i0} and $Y_{s(i),0}$). Two schools with the same initial nonvirginity rates would have the same distribution of ω (controlling for individual characteristics). If, in a later period, Y_{it} is higher or lower, virgins would be more or less likely to search because of the conformity effect, and so they would be more or less likely to become nonvirgins. Similarly, if $Y_{s(i),t}$ is higher or lower, the arrival rate of partners would be greater or lesser. Thus, because the initial variation in Y_0 is absorbed by the types distribution, it is the unpredicted, random innovations in peer and supply group nonvirginity rates that identify the effects of peer norms and partner availability. Put more intuitively, with this approach the estimates of endogenous social effects are based on the magnification of small differences over time in the outcomes, given similar initial conditions.

## 5. ESTIMATES AND COUNTERFACTUAL SIMULATIONS

The estimated search and matching model fits the observed patterns in sexual initiation, and it finds meaningful differences between the two mechanisms of peer norms and partner availability. In what follows, I first describe the estimation results and then use three counterfactual simulations to illustrate the differential impacts of these mechanisms, as well as to forecast the possible results of relevant interventions.

Figure 2 demonstrates the fit of the preferred specification.[45] The observed growth of nonvirginity rates for a synthetic cohort ("8–11 Observed") is plotted alongside predictions from the model ("9 Predicted"). The predicted line shows the rates for the ninth grade cohort projected through the end of twelfth grade. This prediction is formed by starting with the observed virginity statuses in the initial time period (the summer before the ninth-grade cohort entered high school, 1994Q3), and then simulating outcomes for all cohorts going forward. Thus, by the time the ninth-grade cohort reaches the end of high school, the prediction is 15 periods out from the observed data. This is an out-of-sample prediction in the sense that the ninth-grade cohort is only observed through

---

[45]For reasons explained below, the preferred specification uses the second version of (6), the types distribution, which includes individual characteristics (x_i) but not group means ($\bar{x}_m$).
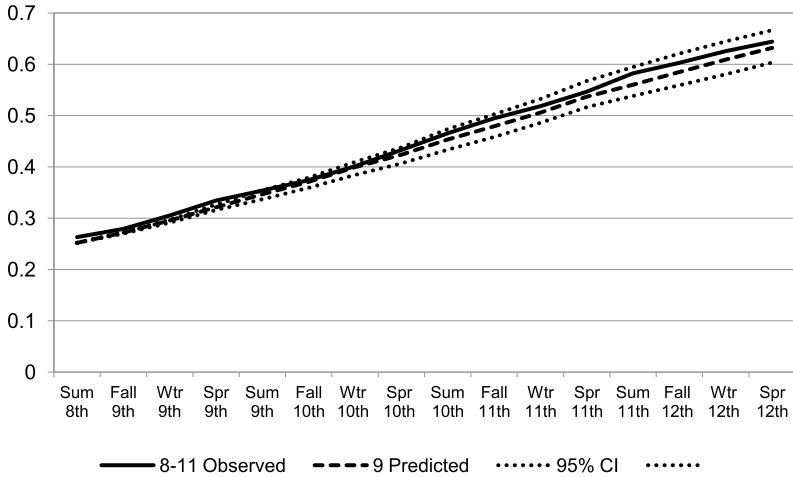
(a) Nonvirginity rates, boys
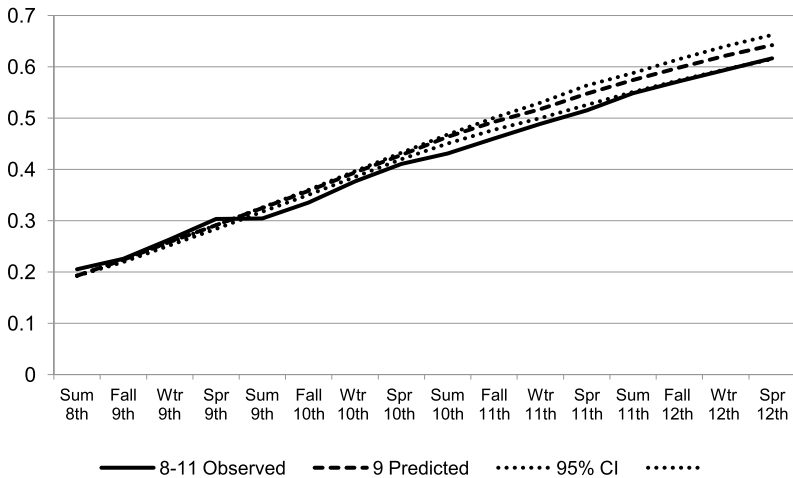
(b) Nonvirginity rates, girls

FIGURE 2. Model fit. "8–11 Observed" is the observed rates for the 8th–11th grade cohorts, combined into a synthetic cohort; "9 Predicted" is the prediction for the 9th grade cohort from the estimated model.

tenth grade, so their predicted outcomes in the eleventh and twelfth grades are shown against the observed outcomes of older cohorts. Confidence intervals are constructed by using a parametric bootstrap of the structural parameters and resimulating outcomes with each draw of the parameter vector. These intervals are relatively narrow compared with the total growth of nonvirginity rates during high school.

The structural parameter estimates from the preferred specification and their standard errors are shown in Table 4. The effect of lagged peer nonvirginity rates on expected utility, γ, is large and significant for both boys and girls. This indicates that peer norms have a substantial influence on the desire to initiate sex. The effect of opposite-gender search behavior on the arrival rate is given by the parameters $\lambda_1 = (\lambda_{11}, \lambda_{12}, \lambda_{13})$. There

TABLE 4. Structural parameter estimates.

Parameter	Boys	Girls
Age		
α	0.082	0.166
	(0.063)	(0.044)
Peer preference interaction		
γ	0.182	0.200
	(0.070)	(0.057)
Arrival rate		
λ_0: 9th grade	−2.612	−2.369
	(0.317)	(0.245)
λ_0: 10th grade	−2.873	−2.420
	(0.308)	(0.245)
λ_0: 11th grade	−2.921	−2.442
	(0.333)	(0.240)
λ_0: 12th grade	−2.865	−2.395
	(0.345)	(0.247)
λ_{11}: same grade	0.556	0.210
	(0.398)	(0.279)
λ_{12}: below/above	0.260	0.022
(boys/girls)	(0.254)	(0.157)
λ_{13}: above/2 above	0.254	0.197
(boys/girls)	(0.183)	(0.154)
Chi-square test of $(\lambda_{11}, \lambda_{12}, \lambda_{13})$		
test statistic	6.99	2.32
(*p*-value)	(0.072)	(0.509)
Type values		
ω^L	−0.270	−0.287
	(0.089)	(0.053)
ω^H	−0.107	−0.089
	(0.066)	(0.056)
Terminal values		
$\nu(\omega^L)$	−1.608	−0.156
	(0.806)	(0.443)
$\nu(\omega^H)$	−0.142	0.722
	(0.880)	(0.938)

(*Continues*)

is one parameter for each of the three grades that provide the endogenous supply of partners for an individual, as defined in Section 3. These three parameters are jointly significant for boys but not for girls. (This difference is discussed in detail with the second counterfactual below.) The age parameter, α, is twice as large for girls compared with boys, which suggests that girls are more influenced by individual development. Finally, the model has two preference types, which I refer to as low and high types, where $\omega^L < \omega^H$. These two types are sufficient to fit the observed growth of nonvirginity rates

TABLE 4. *Continued.*

Parameter	Boys	Girls
Type probabilities (π^H)		
Constant term	0.625	−0.369
	(0.857)	(0.531)
Y_0: 9th grade own gender	0.596	1.765
	(2.438)	(1.635)
Y_0: 9th grade opposite gender	−0.291	2.633
	(2.247)	(1.543)
Black	3.023	0.369
	(2.219)	(0.499)
Younger child	0.741	−0.209
	(0.603)	(0.379)
Only child	1.970	2.908
	(1.494)	(1.249)
Parent educ.	−2.237	−2.386
	(1.432)	(1.125)

TABLE 5. Parameter estimates in alternative specifications.

	Boys				Girls			
	Var in Types Dist. (z)			Alt. Match	Var in Types Dist. (z)			Alt. Match
Parameter	($\cdot$) (1)	x_i (2)	$x_i, \bar{x}_m$ (3)	(4)	($\cdot$) (5)	x_i (6)	$x_i, \bar{x}_m$ (7)	(8)
α (age)	0.082	0.082	0.063	0.098	0.167	0.166	0.166	0.173
γ (peer norms)	0.203	0.182	0.227	0.224	0.204	0.200	0.213	0.213
λ_{11} (availability)	0.689	0.556	0.782	−0.007	0.252	0.210	0.185	0.061
λ_{12} (availability)	0.202	0.260	0.104	−0.003	−0.003	0.022	0.051	−0.080
λ_{13} (availability)	0.182	0.254	0.171	0.049	0.197	0.197	0.166	0.069
Log likelihood	10,116	10,066	10,053	10,070	10,116	10,066	10,053	10,070

during high school, as well as differences in these growth trends along various observable characteristics.[46]

Table 5 compares the key parameter estimates among the alternative specifications (the full set of parameters and their standard errors are listed in Table A.4 in Appendix A). There is little difference in the estimates of the effects of age, peer norms, and partner availability across columns 1–3 and 5–7, which use the three alternative versions of the types distribution ((6)). As with the exercise in Section 3.1, this suggests that the initial nonvirginity rates (Y_0) capture most of the relevant unobserved heterogeneity across schools. Inclusion of the group means in the types distribution (columns 3 and 7)

---

[46]Arcidiacono, Khwaja, and Ouyang (2012) also find that two types are sufficient in their work on adolescent sexual behavior. The growth trends conditional on other observables are available on request.

changes the estimates of the key parameters very little. However, the estimates of the parameters of the types distribution have extremely large and divergent values for boys in this specification (column 2 in Table A.4), which indicates a problem related to the high correlation between individual characteristics and their group averages. As a result, the preferred specification of the model uses the second version of (6), which includes the individual characteristics for additional variation but does not suffer from this problem. Columns 4 and 8 report the results of using the alternative matching technology, which yields similar estimates for γ but has λ_1 essentially equal to zero. This could be interpreted as supporting the main matching technology, where there is no constraint on the number of partners each period, over the alternative, where matches must be one-to-one. However, there are other explanations such as greater measurement error in the ratio of searchers between the two genders compared with the proportion of searchers within one gender. Notably there is little exogenous variation in the gender ratio across schools, which would have helped to generate variation in the endogenous ratio of searchers.[47]

To interpret the values of γ and λ_1, Tables 6 and 7 present average search probabilities and arrival rates by gender and grade, as well as the marginal effects related to these parameters, based on the preferred specification. Table 6 shows that the average probability of search is fairly small among low-type boys and girls in the ninth grade (under 0.2) but increases throughout high school to about 0.5 in the twelfth grade. Among high-type boys and girls, the search probability is already at 0.47 and 0.65 in the ninth grade, respectively, and it rises by about 0.2 more by the twelfth grade. The marginal effects of lagged peer nonvirginity rates on the search decision are substantial in relation to the average search probabilities, especially in younger grades. In the ninth grade, they imply that a 1-standard-deviation increase in the nonvirginity rate among peers would increase the probability that a virgin decides to search by 0.055 for either boys or girls.[48] This is 17% (14%) of the average search probability among boys (girls) in that grade.

TABLE 6. Probability of search among virgins, by type, and marginal effects of lagged peer nonvirginity rates.

	Boys				Girls			
	Probability of Search			Marginal Peer Effect	Probability of Search			Marginal Peer Effect
Grade	Low Type	High Type	Weighted Average		Low Type	High Type	Weighted Average	
9th	0.133	0.466	0.316	0.324	0.196	0.651	0.388	0.341
10th	0.236	0.572	0.415	0.304	0.338	0.787	0.511	0.329
11th	0.366	0.647	0.508	0.222	0.459	0.863	0.601	0.220
12th	0.511	0.675	0.593	0.123	0.506	0.886	0.629	0.119

Note: Weighted averages use the probability that an individual is each type as the weights. Marginal peer effects show the weighted averages of the marginal effects for each type.

[47] The average proportion of male students in each high school in the sample is 48.5%, with a standard deviation of only 3.2%. The interquartile range is from 46.2 to 51.7%.

[48] The standard deviations of peer nonvirginity rates are 0.17 for boys and 0.16 for girls in ninth grade.

TABLE 7. Average arrival rates, and marginal effects of search behavior among the opposite gender.

	Boys				Girls			
		Marginal Supply Effects					Marginal Supply Effects		
	Average Arrival Rate	Same Grade	Grade Below	Grade Above	Average Arrival Rate	Same Grade	One Grade Above	Two Grades Above
Grade								
9th	0.103	0.051	NA	0.024	0.108	0.020	0.002	0.019
10th	0.101	0.050	0.024	0.023	0.106	0.020	0.002	0.019
11th	0.106	0.052	0.025	0.024	0.094	0.018	0.002	NA
12th	0.099	0.050	0.023	NA	0.097	0.018	NA	NA

Table 7 shows that the arrival rate of partners is similar for boys and girls, with an average of about 0.1, which corresponds to an expected wait of 2.5 years. The marginal effects here are smaller in magnitude than those in Table 6, but for boys the combined effect of search behavior among the supply groups can be large. (For girls, the underlying parameters are not jointly significant, so any apparent marginal effects may only reflect sampling noise.) For example, a 1-standard-deviation increase in the search behavior in each of the supply groups raises the arrival rate of partners for a tenth grade boy by 0.02, or 18% of the average rate. It is possible to draw a loose comparison between these marginal effects and those in the matching model of Arcidiacono, Beauchamp, and McElroy (forthcoming). The comparison is difficult because their model is static while mine is dynamic; still, one can make an attempt by accumulating the relevant marginal effects over several periods in my model. The result of this calculation is 0.096 for boys, which is quite similar to the marginal effect of the ratio of searchers on the static matching probability in their model, which I compute to be 0.089 for boys.[49]

Next, to interpret the types distribution, Table 8 shows the probability of being high type (among virgins at the beginning of ninth grade) and the average partial effects of the individual characteristics that condition this probability.[50] Roughly half of the population is high type, with a larger fraction among boys than girls. The partial effects of the individual characteristics are qualitatively similar to the coefficients estimated on these variables in simple hazard models (Tables A.1–A.3 in Appendix A), including the fact that black race has a large partial effect for boys but not girls. Because the preference types include the effects of unobserved location-based factors, however, the relationship between these variables and the desire for sex should not be interpreted as

[49]For their model, I calculate the marginal effect of a change in the ratio of male to female searchers on the matching probability expressed in equation 5 of their paper (December 2012 working paper version). I use the parameters reported in Table 8 of their paper and a one-to-one ratio of searchers, which yields the marginal effect of 0.089. To compare this with my model, I use the probability of finding a partner over eight quarters (i.e., two years, which is close to the median tenure among all students in high school) with a constant arrival rate λ. This probability is $1 - (1 - \lambda)^8$, so the marginal effect of a change in arrival rate is $8(1 - \lambda)^7 \lambda'$. I use $\lambda = 0.1$ and $\lambda' = 0.025$ as rough averages of the estimates for boys reported in Table 7, and the result is 0.096.

[50]These estimates use individuals who are observed when they enter high school (i.e., the eighth- and ninth-grade cohorts). The partial effects are calculated by averaging the individual-level effects.

TABLE 8. Probability of high type and partial effects of individual characteristics.

	Boys	Girls
Probability of high type	0.55	0.43
(among virgins at the beginning of ninth grade)		
Partial effects of		
Black race	0.42	0.06
Being a younger child	0.12	−0.03
Being an only child	0.30	0.47
Parent with 16+ years educ.	−0.41	−0.36

reflecting individual-based factors alone. Particularly in the case of race, the individual characteristics are not randomly distributed across locations, so their relationships with the probability of the high type may reflect some common factors that are not fully captured with the initial nonvirginity rates used to condition the types distribution.

Finally, the parameters for the approximation of equilibrium beliefs ((5)) are reported in Appendix A (Table A.5). These were recovered from the observed nonvirginity rates and then used to estimate the model (columns labeled "Observed"). After estimation of the structural parameters, I then reestimated the approximation with data simulated from the model to check that the specification in (5) is consistent with the model (columns labeled "Simulated"). The new coefficients are generally quite close to those in the original approximation. As a measure of distance between the two, I compute a chi-squared statistic for their difference, using seemingly unrelated regressions. This accounts for sampling variation in the regressions and the naive correlation between the observed and simulated data, but not any further variation in the coefficients due to the estimation of the structural parameters. The statistic is relatively small as reflected by its p-value of 0.75, which indicates that the two sets of coefficients are close in some sense.

5.1 *Counterfactual simulations*

Three counterfactual simulations are presented in Figures 3–5. The first two illustrate the cumulative impacts of school-based peer norms and partner availability by showing what happens when one of these mechanisms is shut down. The third counterfactual forecasts the possible impact of a policy isolating the ninth grade from the rest of high school. In each figure, there are two graphs with the outcomes for boys and for girls, and each graph shows two projections for the ninth-grade cohort: the first uses the estimated model exactly as in Figure 2 ("9 Baseline"); the second uses the model with a modification to yield the simulation ("9 Simulated").[51] Confidence intervals are constructed for the latter using a parametric bootstrap of the structural parameters, as before. For reference, the gray line in these graphs ("8–11 Observed") shows the observed nonvirginity

---

[51]In each counterfactual, equilibrium beliefs must be revised so as to be consistent with the modified model. I do this by estimating the approximation ψ on data simulated from the modified model, and then simulating new data based on the new beliefs. I repeat this process until the parameters of the beliefs approximation converge, which occurs in fewer than 10 iterations.
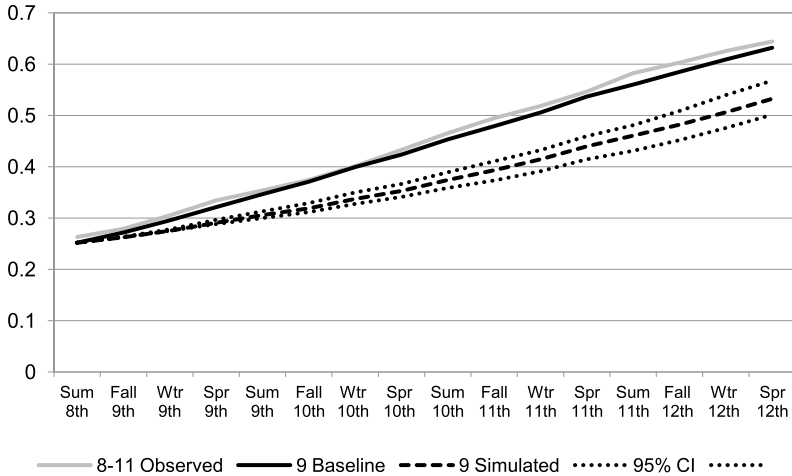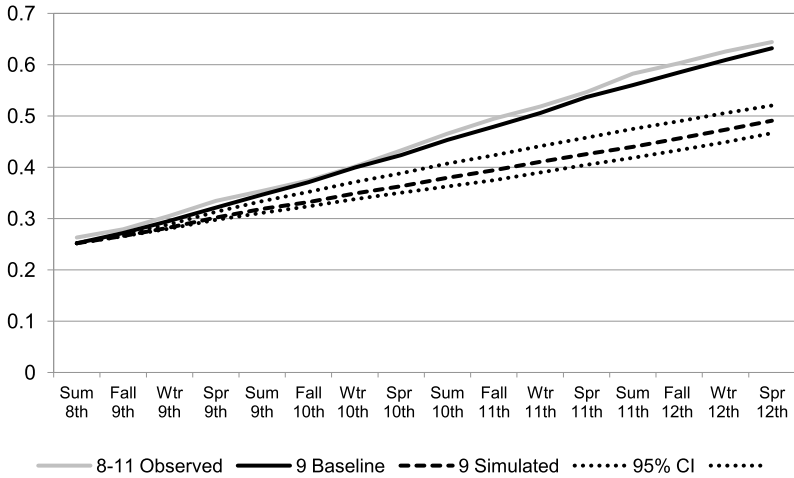
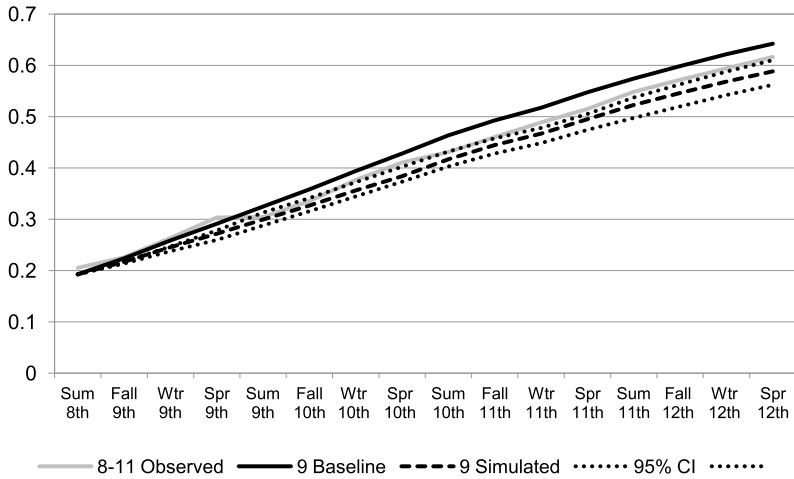(a) Nonvirginity rates, boys
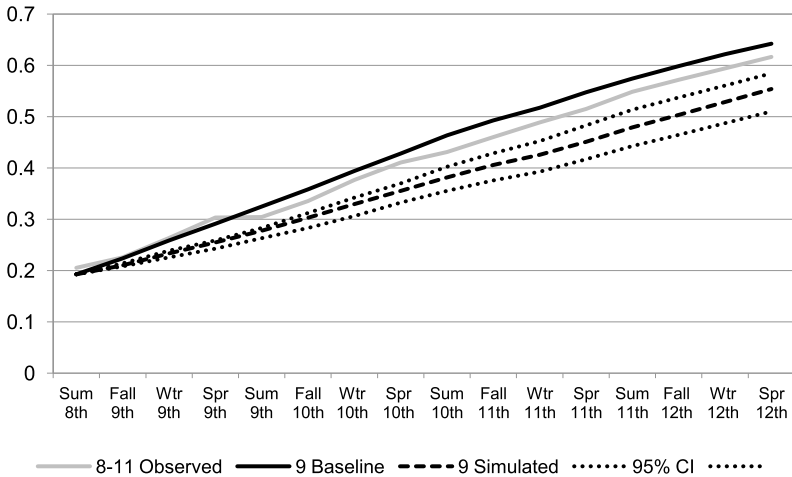
(b) Nonvirginity rates, girls

FIGURE 3. Eliminated effect of peer norms on search decisions. "8–11 Observed" is 8th–11th grade cohorts, combined; "9 Baseline" is the prediction for the 9th grade cohort from estimated model; "9 Simulated" is the prediction from modified model.

rates for the synthetic cohort, but the relevant comparisons are between the estimated model and the simulations.

To demonstrate the cumulative impact of peer-group norms, the simulation in Figure 3 removes their influence by setting the parameters $\gamma_b = \gamma_g = 0$. The result is that the fraction of boys (girls) who become sexually active during high school is reduced by 0.10 (0.09), which is 26% (20%) of the total fraction who initiate during high school (0.38 for boys and 0.45 for girls). The 95% confidence interval for the amount of this reduction goes from 0.06 to 0.13 (also 0.06 to 0.13 for girls), that is, 17–34% (13–30%) of the total. The relative impact of peer norms is even larger in younger grades: the number of individuals who initiate sex in the ninth or tenth grade falls by 41% for boys and 31%
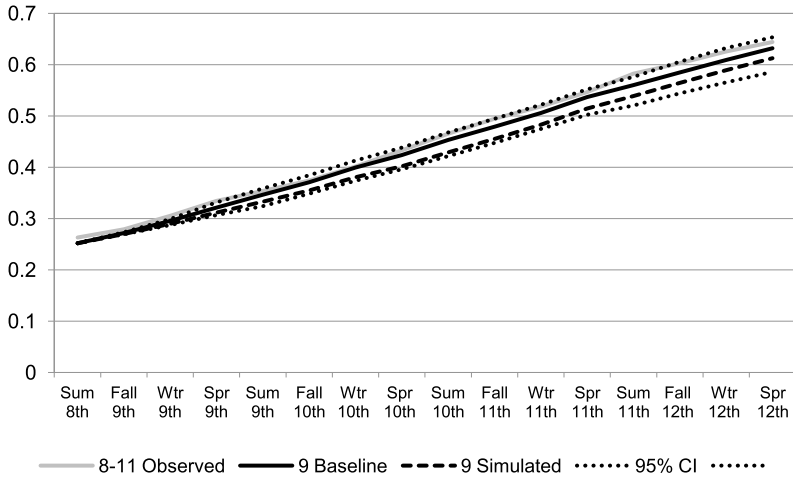
(a) Nonvirginity rates, boys
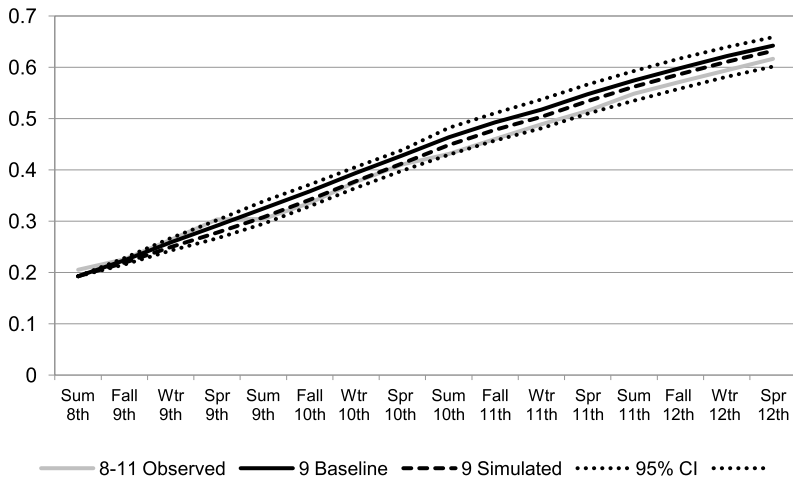

(b) Nonvirginity rates, girls

FIGURE 4. Eliminated effect of opposite-gender search behavior on arrival rates. "8–11 Observed" is 8th–11th grade cohorts, combined; "9 Baseline" is the prediction for the 9th grade cohort from estimated model; "9 Simulated" is the prediction from modified model.

for girls (95% confidence interval: 33–48% for boys, 25–41% for girls). Overall, these results indicate that peer norms have a substantial impact on sexual initiation during high school. This suggests that interventions targeting social norms could be effective if they can actually inhibit this mechanism to some extent.

These effects of peer norms can be loosely compared to estimates of the influence of same-gender best friends on sexual initiation in Card and Giuliano (2013). Their model does not differentiate among mechanisms, but it seems reasonable to assume that this effect is more related to norms than supply. They estimate that the probability of becoming sexually active over the course of a year increases by one third if the best friend does as well. This is roughly similar to the 20–25% drops in initiation rates above, and

(a) Nonvirginity rates, boys

(b) Nonvirginity rates, girls

FIGURE 5. Remove ninth grade from high school. "8–11 Observed" is 8th–11th grade cohorts, combined; "9 Baseline" is the prediction for the 9th grade cohort from estimated model; "9 Simulated" is the prediction from modified model.

it is not surprising to see somewhat larger effects from best friends compared with all same-gender classmates.

The simulation in Figure 4 eliminates the effect of partner availability at school by setting the parameters $\lambda_{b1} = \lambda_{g1} = 0$. This means that the arrival rate does not depend on opposite-gender search behavior in the local market; however, the arrival rate is still nonzero due to the presence of the external market, so virgins can still find partners and become sexually active. The results indicate that the availability of boys at school has very little effect on the initiation rate for girls, while the availability of girls at school impacts boys substantially. Without any girls at their schools who are looking for sexual partners (and without any compensating behavior), the fraction of boys who become

sexually active during high school falls by 0.14, which is 37% of the total. The reduction for girls is much smaller with a point estimate of 0.05, that is, 12% of the total. The confidence intervals for these reductions are reasonably narrow, with a width of 0.05 at the end for both boys and girls.[52]

Two explanations for this difference between boys and girls would be consistent with the model. First, girls may be more likely to find partners in the external market. In the model, the parameter λ_0 is intended to capture the probability of finding a partner in the external market, and it is indeed larger for girls than for boys (Table 4).[53] A second explanation is that boys may be more likely to have multiple partners. Either way, the proportion of boys searching in the local market would matter less to girls.

We can find some evidence related to these explanations in the detailed relationship histories in Add Health, which were taken on up to three romantic partners. Among other information, these histories include whether each partner was from the respondent's school, and the first and last dates of sexual intercourse. Based on these data, Table 9 reports the fraction of first sexual partners who went to the same school, and the fraction of second sexual relationships that apparently began before the first ended. Girls are more likely to have their first partner come from outside their school (56% for girls vs. 44% for boys), whether from a different school or someone who is no longer in school. On the other hand, having multiple partners does not appear to be common for either boys or girls. Only 5% of each gender report a second sexual relationship that began before their first ended.[54] Based on this evidence, it appears that the proportion

TABLE 9. Additional information on sexual relationships.

	Boys		Girls
First partner was from		(%)	
Same school	56		44
Different school	41		35
Not in school	3		21
Second relationship began		(%)	
After first ended	13		13
Before first ended	5		5
No second rel. reported	82		82
N	1686		2123

Note: Sample is restricted to individuals with month and year reported for first partner in the relationship histories.

[52]One might think that the confidence interval for girls in this simulation should cover the baseline projection, because λ_{g1} is not statistically significant. However, this simulation sets both λ_{b1} and λ_{g1} to zero, and the confidence intervals reflect the cumulative influence of these effects over time.

[53]The implied arrival rate of match offers from the external market is about 0.083 ($= \Lambda(-2.4)$) per quarter for girls and 0.057 ($= \Lambda(-2.8)$) for boys.

[54]Note that while the fraction of individuals with overlapping relationships in this sample is small (5%), among individuals having second partners, the proportion with overlapping relationships is over one-quarter (5%/18%). This supports the main specification of the matching technology which allows for multiple concurrent partners.

of partners found in the external market could explain much of the difference between boys and girls in the effect of partner availability at school, but not the frequency of having multiple partners. There could be other explanations for the relatively small value of λ_1 for girls, of course, such as a lack of variation in the proportion of searchers among the opposite gender or some misspecification in the matching technology. However, any such explanation would have to involve differences between boys and girls so as to generate the rather different values of λ_{b1} and λ_{g1}.

The simulation eliminating the effect of partner availability could be interpreted as an upper bound on the potential for single-sex schools to delay sexual initiation. The results suggest that boys would be much more impacted than girls. However, to the extent that boys would compensate by increasing their search effort in the external market (a choice that is outside the model), the effect of such a policy would be reduced. Figure 5 presents the results of a more modest and perhaps more realistic policy simulation, which involves isolating the ninth grade from the rest of high school. In the Add Health data, 5% of ninth graders are in a middle school or otherwise separated from the older grades in high school. The counterfactual is accomplished by setting the arrival rate parameters that apply to the older supply groups for ninth graders (i.e., λ_{g12}, λ_{g13}, and λ_{b13}) and the younger supply group for tenth grade boys (λ_{b12}) to zero. The decrease in sexual initiation in the ninth grade is about 14% for both boys and girls. However, this reduction dissipates rapidly and the confidence intervals for the simulation always include the baseline projection, so we cannot reject the possibility that this policy would have no impact.

6. Conclusion

This work estimates a search and matching model for sexual partners in high school, so as to recover the effects of two social mechanisms that separately influence demand and supply in this market. These mechanisms—social norms among peers and partner availability at school—would be affected by distinct interventions: for example, educational programs about peer pressure for the former and gender segregation for the latter. The model is flexibly specified so that separate effects can be recovered for each mechanism for boys and for girls. The results indicate that peer-group norms have a large effect on the timing of sexual initiation for both genders while the availability of partners at school is only important for boys. Among other implications, this broadly suggests that gender segregation would not have a large impact on pregnancy rates among teenage girls, which is one of the chief concerns that motivates policy interest in adolescent sexual activity.[55]

Beyond this application, there are many areas where behavior may be governed more by social influence than by traditional, monetized market interactions. A large literature in economics has developed over the past 20 years that studies these phenomena. Much of the empirical literature estimates some kind of a composite effect of social

---

[55]In the 1995 State of the Union address, President Bill Clinton declared the "epidemic" of teenage pregnancy and out-of-wedlock childbearing to be "our most serious social problem." (Transcript downloaded from *The American Presidency Project* website, http://www.presidency.ucsb.edu/ws/index.php?pid=51634, accessed 11/10/09.)

interactions, which does not distinguish among the mechanisms that may drive behavior in a particular application. However, recently a growing number of authors have attempted to examine the mechanisms behind endogenous social effects. Relative to that work, this paper offers an approach to recover the effects of the underlying social mechanisms by directly specifying them as part of a structural economic model. Such an approach may be useful for work in other contexts when assessing specific mechanisms is important for understanding behavior and making policy recommendations.

References

Aguirregabiria, V. and P. Mira (2007), "Sequential estimation of dynamic discrete games." *Econometrica*, 75 (1), 1–53. [665, 674]

Arcidiacono, P., A. Beauchamp, and M. McElroy (forthcoming), "Terms of endearment: An equilibrium model of sex and matching." *Quantitative Economics*. [665, 666, 692]

Arcidiacono, P., A. Khwaja, and L. Ouyang (2012), "Habit persistence and teen sex: Could increased access to contraception have unintended consequences for teen pregnancies?" *Journal of Business & Economic Statistics*, 30 (2), 312–325. [690]

Argys, L. M., D. I. Rees, S. L. Averett, and B. Witoonchart (2006), "Birth order and risky adolescent behavior." *Economic Inquiry*, 44 (2), 215–233. [677]

Bajari, P., C. L. Benkard, and J. Levin (2007), "Estimating dynamic models of imperfect competition." *Econometrica*, 75 (5), 1331–1370. [665, 674]

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011), "Identification of social interactions." In *Handbook of Social Economics*, Vol. 1 (J. Benhabib, A. Bisin, and M. O. Jackson, eds.), 853–964, Elsevier, Amsterdam. [667, 685]

Brock, W. A. and S. N. Durlauf (2001a), "Discrete choice with social interactions." *Review of Economic Studies*, 68 (2), 235–260. [670, 685]

Brock, W. A. and S. N. Durlauf (2001b), "Interactions-based models." In *Handbook of Econometrics*, Vol. 5 (J. J. Heckman and E. E. Leamer, eds.), 3297–3380, Elsevier, Amsterdam. [665, 666, 685, 686]

Card, D. and L. Giuliano (2013), "Peer effects and multiple equilibria in the risky behavior of friends." *Review of Economics and Statistics*, 95 (40), 1130–1149. [666, 695]

Case, A. C. and L. F. Katz (1991), "The company you keep: The effects of family and neighborhood on disadvantaged youths." Working Paper 3705, National Bureau of Economic Research. [666]

Centers for Disease Control and Prevention (CDC) (2008), "Youth risk behavior surveillance: United States, 2007." *Morbidity and Mortality Weekly Report*, 57 (SS-4), 1–131. [663]

Choo, E. and A. Siow (2006), "Who marries whom and why." *Journal of Political Economy*, 114 (1), 175–201. [665]

Clark, A. E. and Y. Lohéac (2007), "'It wasn't me, it was them!' Social influence in risky behavior by adolescents." *Journal of Health Economics*, 26 (4), 763–784. [666]

De Giorgi, G. and M. Pellizzari (2014), "Understanding social interactions: Evidence from the classroom." *The Economic Journal*, 124 (579), 917–953. [667]

de Paula, Á. (2009), "Inference in a synchronization game with social interactions." *Journal of Econometrics*, 148 (1), 56–71. [666]

Drewianka, S. (2003), "Estimating social effects in matching markets: Externalities in spousal search." *Review of Economics and Statistics*, 85 (2), 409–423. [667]

Duflo, E., P. Dupas, and M. Kremer (2011), "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya." *American Economic Review*, 101 (5), 1739–1774. [667]

Eckstein, Z. and K. I. Wolpin (1990), "Estimating a market equilibrium search model from panel data on individuals." *Econometrica*, 58 (4), 783–808. [665]

Evans, W. N., W. E. Oates, and R. M. Schwab (1992), "Measuring peer group effects: A study of teenage behavior." *Journal of Political Economy*, 100 (5), 966–991. [666]

Fletcher, J. M. (2007), "Social multipliers in sexual initiation decisions among U.S. high school students." *Demography*, 44 (2), 373–388. [666]

Fruehwirth, J. C. (2014), "Can achievement peer effect estimates inform policy? A view from inside the black box." *Review of Economics and Statistics*, 96 (3), 514–523. [667]

Gaviria, A. and S. Raphael (2001), "School-based peer effects and juvenile behavior." *Review of Economics and Statistics*, 83 (2), 257–268. [666]

Gourieroux, C. and A. Monfort (1997), *Simulation-Based Econometric Methods.* Oxford University Press, Oxford. [683]

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003), "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18 (5), 527–544. [686]

Hotz, V. J. and R. A. Miller (1993), "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies*, 60 (3), 497–529. [665]

Hoxby, C. (2000), "Peer effects in the classroom: Learning from gender and race variation." Working Paper 7867, National Bureau of Economic Research. [686]

Keane, M. P. and K. I. Wolpin (1994), "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence." *Review of Economics and Statistics*, 76 (4), 648–672. [674]

Keane, M. P. and K. I. Wolpin (1997), "The career decisions of young men." *Journal of Political Economy*, 105 (3), 473–522. [685]

Kinsman, S. B., D. Romer, F. F. Furstenberg, and D. F. Schwarz (1998), "Early sexual initiation: The role of peer norms." *Pediatrics*, 102 (5), 1185–1192. [670]

Krusell, P. and A. A. Smith (1998), "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106 (5), 867–896. [672, 674]

Lee, D. and K. I. Wolpin (2006), "Intersectoral labor mobility and the growth of the service sector." *Econometrica*, 74 (1), 1–46. [672, 674, 675]

Levine, P. B. (2001), "The sexual activity and birth control use of American teenagers." In *Risky Behavior Among Youths: An Economic Analysis* (J. Gruber, ed.), University of Chicago Press, Chicago. [677]

Lundborg, P. (2006), "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25 (2), 214–233. [666]

Manlove, J., A. Romano-Papillo, and E. Ikramullah (2004), "Not yet: Programs to delay first sex among teens." National Campaign to Prevent Teen Pregnancy. [664]

Manski, C. F. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60 (3), 531–542. [665, 666, 685, 686]

Mas, A. and E. Moretti (2009), "Peers at work." *American Economic Review*, 99 (1), 112–145. [667]

Moffitt, R. A. (2001), "Policy interventions, low-level equilibria, and social interactions." In *Social Dynamics* (S. N. Durlauf and H. Peyton Young, eds.), Economic Learning and Social Evolution, Vol. 4, 45–82, Brookings Institution Press, Washington, DC; MIT Press, Cambridge. [665]

Mortensen, D. T. and C. A. Pissarides (1994), "Job creation and job destruction in the theory of unemployment." *Review of Economic Studies*, 61 (3), 397–415. [670]

Murphy, K. M. and R. H. Topel (2002), "Estimation and inference in two-step econometric models." *Journal of Business & Economic Statistics*, 20 (1), 88–97. [683]

Nakajima, R. (2007), "Measuring peer effects on youth smoking behaviour." *Review of Economic Studies*, 74 (3), 897–935. [666, 667]

National Research Council, Panel on Adolescent Pregnancy and Childbearing (1987), *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*. National Academy of Sciences, Washington, DC. [670, 677]

Oettinger, G. S. (1999), "The effects of sex education on teen sexual activity and teen pregnancy." *Journal of Political Economy*, 107 (3), 606–644. [681]

Pakes, A., M. Ostrovsky, and S. Berry (2007), "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)." *RAND Journal of Economics*, 38 (2), 373–399. [665, 674]

Santelli, J. S., J. Kaiser, L. Hirsch, A. Radosh, L. Simkin, and S. Middlestadt (2004), "Initiation of sexual intercourse among middle school adolescents: The influence of psychosocial factors." *Journal of Adolescent Health*, 34 (3), 200–208. [670]

Sieving, R. E., M. E. Eisenberg, S. Pettingell, and C. Skay (2006), "Friends' influence on adolescents' first sexual intercourse." *Perspectives on Sexual and Reproductive Health*, 38 (1), 13–19. [670]

Sirakaya, S. (2006), "Recidivism and social interactions." *Journal of the American Statistical Association*, 101 (475), 863–877. [666]

Trogdon, J. G., J. Nonnemaker, and J. Pais (2008), "Peer effects in adolescent overweight." *Journal of Health Economics*, 27 (5), 1388–1399. [666]

Widmer, E. D. (1997), "Influence of older siblings on initiation of sexual intercourse." *Journal of Marriage and Family*, 59 (4), 928–938. [677]