Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 12-015

"Peer Effects in Sexual Initiation: Separating Demand and Supply Mechanisms"

by

Seth Richards-Shubik

Peer Effects in Sexual Initiation:
Separating Demand and Supply Mechanisms*

Seth Richards-Shubik[†]
Carnegie Mellon University

January 17, 2012

Abstract

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. The existence of a composite effect of social interactions is also confirmed using a standard method: instrumental variables estimation of linear regressions.

*I thank my thesis advisors for their dedication, support, and hours of their time: Kenneth I. Wolpin, David M. Cutler, Elena Krasnokutskaya, and Áureo de Paula. I am also grateful to John Cawley, Guy David, Steven N. Durlauf, Sara Markowitz, Ellen R. Meara, and Nirav Mehta for helpful comments, and to the National Bureau of Economic Research for computer facilities. This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

[†]Email: sethrs@andrew.cmu.edu

1 Introduction

About one-half of the students in grades nine through twelve in the United States are sexually experienced (CDC 2008a). The prevalence of sexual activity among adolescents raises substantial concerns, largely because of associated risks such as unplanned pregnancy and sexually transmitted disease. And as with many risky behaviors, peers are often pointed to as a major influence on the decisions adolescents make about sex.

A rapidly growing body of research in economics examines peer effects. However the emphasis of empirical work in this area is on the identification of simple statistical models of behavior, with little attention to clearly defined and theoretically motivated mechanisms.[1] These models define an endogenous peer effect as the change in the probability of an outcome for an individual caused by a change in the distribution of that outcome (usually the mean) among some reference group. We can think of this as the *composite* effect of social interactions, and clearly many mechanisms could be behind such an effect.

Many valuable contributions have come from the study of this composite effect. However the usual econometric models, where parameters are defined purely in terms of observed behavioral relationships, are subject to the Lucas Critique (Lucas 1976). There is no reason to believe their predictions will be valid under alternative policy regimes. In fact, this point was borne out in a recent policy experiment at the U.S. Air Force Academy (Carrell, Sacerdote, and West 2011). Incoming students were assigned to squadrons in a way that was predicted to maximize average academic performance, based on peer effects estimated in prior years. Instead, the performance of the targeted group declined.[2]

[1] In their recent *Handbook* chapter on social interactions, Blume et al. (2011) write: "A final area that warrants far more research is the microfoundations of social interactions. 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)

[2] Cooley (2010) provides some theoretical insight into this perverse result. If performance spillovers result from unobserved ability or effort, the observed relationship between exogenous peer characteristics and individual performance can have the opposite sign of the true relationship, when peer performance is also

To date, a very limited number of empirical studies have examined the mechanisms behind peer effects, and they have done so informally or indirectly. Mas and Moretti (2009) present the most compelling informal evidence, using data on supermarket cashiers. They show that productivity spillovers from faster to slower cashiers occur only when the faster ones are positioned where they can see their slower coworkers. They interpret this as evidence that the mechanism for the spillover is social pressure (e.g., informal sanctions) rather than prosocial behavior (e.g., enjoying cooperation). Lavy and Schlosser (2011) study mechanisms behind the effect of classroom gender ratios on academic performance. To do so, they substitute various measures of student perceptions about their classroom social environment in place of academic performance as the dependent variable in their regression models. If the coefficient on the gender ratio is significant, they interpret this as evidence that the peer effect is "partially being driven by [that] particular mediating factor." However they cannot directly test whether these measures in fact mediate the relationship between gender ratios and academic performance, and they acknowledge that this rich analysis provides only suggestive evidence on possible mechanisms.[3]

In this paper, I present a formal empirical analysis of two distinct social mechanisms in the onset of sexual activity: social norms among peers and the availability of partners at school. These mechanisms fall separately on the demand side and the supply side of the market for sexual partners. They are also targeted by different interventions. Some interventions to delay sexual initiation include an educational program against peer-group norms that promote sex (Manlove, Romano-Papillo, and Ikramullah 2004). Such programs target a peer effect on the demand for sexual partners. An alternative is to restrict the supply of partners at school. Single-sex schools represent one way to do this, but a less drastic option is to isolate the ninth grade from the older grades in high school, as in some

included in the estimated model.

[3]As they write, "we cannot identify the causal effect of these mechanisms on outcomes" (p. 21).

districts where the ninth grade is in middle school.

To separate the effects of peer-group norms and partner availability, I estimate an equilibrium model of the market for sexual partners in high school. 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.[4] I use a search and matching framework, so 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. Thus the effect of partner availability is defined as the change in the probability of finding a match due to changes in the search behavior among others at school.

Previous work has estimated a composite effect of school-based social interactions on sexual initiation (Fletcher 2007). However standard methods cannot distinguish between demand and supply side mechanisms in equilibrium, when there are preference interactions. Because peer-group norms operate on both sides of the market, any exogenous change in the demand from one side will indirectly shift the supply curve on the other side as well, due to the change in sexual activity which necessarily involves both sides. Thus careful attention is required to disentangle the effects of peer norms and partner availability, which motivates the formal use of an equilibrium model.

The data for my analysis come from the National Longitudinal Study of Adolescent Health (Add Health), which provides a nationally representative sample of U.S. high school students in the mid-1990s. I follow 14,300 students over two years using retrospective sexual histories taken in two rounds of interviews. The observation of virginity status over time makes it possible to estimate a dynamic search and matching model as in the job search literature (e.g., Eckstein and Wolpin 1990) rather than a static matching model as in recent work on marriage and dating markets (e.g., Choo and Siow 2006; Arcidiacono, Beauchamp,

[4] To be exact, "norms" is a reasonable interpretation for the influence of peer-group nonvirginity rates on individual preferences. See section 2 for evidence on norms from the sociology literature.

and McElroy 2011).

To solve the model and proceed with estimation, I apply an insight from the literature on dynamic discrete games, which is to use observed state transition probabilities directly as rational beliefs (Aguirregabiria and Mira 2007; Bajari, Benkard, and Levin 2007; Pakes, Ostrovsky, and Berry 2007).[5] I adapt this technique to be feasible in my application, where the large (but finite) number of agents produces a complicated distribution of outcomes, by approximating the evolution of nonvirginity rates—the aggregate state—as an autoregressive process. This draws on Krusell and Smith (1998) and Lee and Wolpin (2006), who use simple first-order Markov processes to approximate rational expectations forecasts, and who find that such approximations perform very well. The combination of these two methods, obtaining beliefs directly from the data by using a feasible approximation, greatly simplifies and speeds the computational process because there is no need to solve for a new equilibrium with each set of candidate parameters.

The use of an approximation raises a concern about the "reflection" problem (Manski 1993), a common identification problem in work on social interactions. However I demonstrate that this problem does not apply in my model. Trivially it is because the approximation is nonlinear (Brock and Durlauf (2001b) show that the reflection problem depends on linearity), but I also show that the standard assumption of stable preferences over time provides a key identifying restriction. Two other identification problems in social interactions models are the potential correlation of omitted variables among peers, and selection into peer groups (Moffit 2001; Brock and Durlauf 2001b). To address the correlation issue, I specify a distribution of permanent unobserved heterogeneity that is a function of initial nonvirginity rates when a cohort enters high school, and thus is naturally correlated among peers. To avoid a selection problem, I use gender and grade cohorts to define peer groups,

[5]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.

rather than endogenous social groups like sport teams or nominated friends.[6] A final challenge for identification, more specific to my model, is the separation of the arrival rate from the search probability. I take advantage of data on the arrival of subsequent partners after the first to nonparametrically identify the arrival rate.

Broadly speaking, my identification strategy fits with a growing number of papers that use longitudinal data to identify social interactions.[7] Among these, Hanushek et al. (2003) and Jackson and Bruegmann (2009) use lagged peer outcomes in place of their current outcomes in order to circumvent simultaneity bias and the reflection problem. My use of lags bears some relation to this approach, but expectations about current and future peer outcomes also affect behavior in my model. Mas and Moretti (2009) and Arcidiacono et al. (2011) exploit individual-level panel data as well, although they model peer effects as coming from exogenous fixed effects. Finally, a small number of authors have previously considered social interactions in duration models. Brock and Durlauf (2001b) discuss identification in continuous-time models based on parametric hazard functions, and Sirakaya (2006) estimates such a model applied to recidivism. De Paula (2009) presents a nonparametric test for social interactions in a duration model, based on simultaneous transitions among peers.

The results I obtain indicate that 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 share of individuals who initiate sex during the ninth or tenth grade falls by 0.07 for both boys and girls (a 41% decline for boys and 31% for girls). Changes in the availability of partners at school also impact the initiation rate for boys, but are not statistically significant for girls. The effect on boys is large: for example, for tenth grade boys a one-standard-deviation increase in the share of girls searching for a sexual partner

[6]The selection of school districts is another concern in this literature, but in my model this would be captured by the permanent component of preferences that is correlated within schools.

[7]A related strategy is to use multiple cohorts within schools, which does not require longitudinal data on individuals. This approach was introduced by Hoxby (2000).

raises the probability of finding a partner each period by 18 percent. Overall, these results are consistent with the composite effects which I find using a standard instrumental variables (IV) approach. Previous studies have also found large composite effects of social interactions among adolescents on various risky behaviors, including sexual initiation (Fletcher 2007), 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).

Policy simulations from the estimated model indicate that educational interventions reducing the effect of peer norms could have a larger impact than isolating the ninth grade from older grades to restrict the supply of partners. While the complete elimination of peer influence among adolescents seems unattainable (and perhaps undesirable), partially reducing the peer effect on demand would still have a substantial impact. On the other hand, removing the ninth grade from high school would decrease initiation in that year by about 14 percent for both boys and girls, but the effect dissipates over time.

The rest of this paper is organized as follows. The next section gives further background on teenage sexual activity and summarizes existing evidence on peer and other influences in this behavior. Section 3 presents the model. Section 4 describes the data and presents standard IV estimates of the composite effect of social interactions. Section 5 describes the estimation procedure and contains detailed arguments on identification. Section 6 presents the results and counterfactual simulations.

2 Background on Teenage Sexual Behavior in the U.S.

Data from the Youth Risk Behavior Survey, a national survey of high school students, shows that the fraction individuals in grades nine through twelve who were sexually experienced declined from 54 percent in 1991 to 46 percent in 2001 and then increased (not statistically significantly) to 48 percent in 2007.[8] Sexual activity is highly persistent among teenagers. Only 9 percent of sexually experienced teenage boys and girls report having had sex only once, compared with 69 (76) percent of nonvirgin boys (girls) who report having sex in the past three months and 87 (91) percent in the past year.[9] This persistence indicates the importance of initiation, which Arcidiacono, Khwaja, and Ouyang (2011) measure explicitly as a large fixed-cost-like term in preferences.

While sexual initiation is a normal part of human development, most of the policy interest in adolescent sexual activity focuses on the risks associated with sex. Among these risks, unplanned pregnancy receives the most attention. For example, 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."[10] Teenage childbearing is associated with lower educational attainment for the mothers and an increased probability of incarceration for male children, among other negative outcomes, and there are large public expenditures on the children of teenage mothers (Hoffman 2006). Sexually transmitted diseases (STDs) are another major risk. Teenagers and 20-24 year-olds have the highest prevalence of STDs of any age groups (CDC 2008b). Considering the eight most common STDs, Chesson et al. (2004) estimate that the lifetime medical cost of treating the STDs acquired over one year by 15 to 24 year-olds totals $6.5 billion. Finally, there is work indicating possible direct effects of early sexual initiation on psychological well-being and academic performance. Sabia (2007) and Sabia and Rees (2008) estimate that boys who initiate sex before age 16 then have decreased GPAs and girls who initiate before age 17 are then more likely to have symptoms of depression, controlling for individual fixed effects.

A number of individual-based factors have been studied in relation to sexual initiation

[8] CDC fact sheet on "Trends in the Prevalence of Sexual Behaviors, National YRBS: 1991-2007," `http://www.cdc.gov/HealthyYouth/yrbs/pdf/yrbs07_us_sexual_behaviors_trend.pdf`, accessed 4/28/09.

[9] These figures are calculated from data published in Abma et al. (2004).

[10] Transcript from "The American Presidency Project" website, `http://www.presidency.ucsb.edu/ws/index.php?pid=51634`, accessed 11/10/09.

and sexual activity among adolescents. The National Research Council (NRC) Panel on Adolescent Pregnancy and Childbearing (1987, chapter 4) provides a useful overview of characteristics that predict early initiation. These include black race and low socioeconomic status meaured by parental education and family income, although some recent research finds the association with race is mainly for boys (Levine 2001; Michael and Bickert 2001). Early onset of puberty is another strong and plausibly exogenous predictor, which is well measured in girls as the age of menarche (NRC Panel 1987; Miller et al. 1997). Several authors have studied the relationship between substance use and sexual behavior, with mixed evidence about the effect of alcohol or drug use on sexual activity (Rees, Argys, and Averett 2001; Sen 2002; Markowitz, Kaestner, and Grossman 2005). Finally, 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 hazard rate for initiation among girls.

The role of peer norms in sexual initiation has been documented primarily by work in psychology and sociology. Kinsman et al. (1998), Santelli et al. (2004), and Sieving et al. (2006) measure peer norms through self-reported perceptions about the level of sexual activity among peers, peer attitudes toward sex, or the social gains for becoming sexually active. All three studies find that norms defined in these ways have a significant association with the probability of initiating sex. In addition, earlier work surveyed by the NRC Panel (1987) indicates that peer norms regarding sexual behavior are established within gender for adolescents. As for econometric work, Fletcher (2007) is the only analysis estimating a social multiplier on sexual behavior per se, although Case and Katz (1991) and Evans, Oates, and Schwab (1992) estimate social effects on teenage childbearing.

Some interventions to delay sexual initiation explicitly target peer norms. Two examples are “Safer Choices,” first implemented in 1993 with 2,000 ninth and tenth grade students in California and Texas, and “Draw the Line/Respect the Line,” first implemented in 1997 with 1,500 middle school students in California (Manlove, Romano-Papillo, and Ikramullah 2004).[11] These programs consist of about 20 classroom sessions over two or three school years, with some sessions devoted to learning about social norms and practicing communication skills via role-playing exercises. The existence of interventions such as these demonstrates a specific policy interest in the effect of peer norms, as distinct from the composite effect of social interactions.

3 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, there is a probability of finding a partner per period which depends on the search behavior among virgins and nonvirgins of the opposite gender. Equilibrium is defined within a local market for partners, which is the student body at a high school. There is also an external market, which appears through an exogenous probability of finding a partner from outside the school.

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 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

[11] These programs were evaluated by comparing outcomes across schools that were randomly assigned into treatment or control groups. The results found that Draw the Line reduced sexual initiation among boys by one third but Safer Choices did not have an effect on the population as a whole, although there was a decrease in initiation among Latinos (Manlove et al. 2004).

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. In connection to this, subsequent decisions related to sexual activity are suppressed (e.g., contraceptive use and abortion).[12] Further decisions and additional structure in the payoffs are not necessary for my 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 large portion of the population. And again, it avoids the decisions to have multiple partners or end a relationship.

Finally, match probabilities do not depend on own or partner characteristics. Including them would introduce sorting behavior, which is not the purpose of this research. Thus 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 typical characteristics one would think to use to add heterogeneity to the arrival rate or match probabilities are 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.

3.1 Model Specification

The model applies to individuals, i, located in a local market for partners, m, who have a gender, $\tau \in \{b, g\}$.[13] Virginity status at the start of a period is denoted $y_{i,t-1}$, with 0

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

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

meaning virginity. Each period virgins in the market simultaneously make search decisions $d_{it} \in \{0, 1\}$, and for those who search there is a probability of finding a partner and thereby initiating sex. Nonvirgins always search.

Age, a, is defined socially as quarter within grade in high school. The model covers the fall of ninth grade ($a = 1$) through the spring of twelfth grade ($a = 15 \equiv A$). Time, t, is also measured in quarters, and is needed separately from age to track multiple cohorts at once. However in the exposition, the model is presented from the perspective of a reference cohort for which time and age are equal ($a_{it} = t$). Also, all functions are gender-specific, but gender subscripts are generally suppressed unless needed for clarity.

The probability of finding a partner each period is expressed by the arrival rate, λ_{it}, which is a function of the proportion of searchers among the opposite gender at the school. This proportion is denoted N_{it}, and it includes the behavior of both virgins and nonvirgins. The function is specified with the logistic 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 school, which corresponds to 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.

This arrival rate function does not depend on the search behavior within an individual's own gender, meaning it essentially ignores competition for partners among same-gender individuals. 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 which makes the opposite assumption. The alternative specification uses the ratio of the numbers of opposite-gender to own-gender searchers: for example, $\lambda_{it} = \lambda_{a_{it}}(N^g_{it}/N^b_{it})$ 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 that is commonly used in the macroeconomic labor literature such as Mortensen and Pissarides (2004). To motivate this kind of matching function, it is typically assumed that matches are strictly one-to-one per period.[14]

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 IID mean-zero preference shock, ϵ_{it}, which is private information. Peers are individuals of the same gender in the same grade as i. The per-period utility for a nonvirigin is thus

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

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). To be precise, the term relates lagged peer nonvirginity rates to the expected utility of sex, and I interpret this as an effect of social norms based on the evidence from sociological work on sexual initiation discussed in section 2.[15] 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 captures differences in the desire for sex, as well as differences

[14] 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.

[15] Also, the use of *lagged* peer nonvirginity rates is supported by sociological work such as Kinsman et al. (1998) which specifically asks for perceptions about how many peers are *already* sexually experienced.

in the costs of pregnancy and STDs, or the perceptions of these risks.

Individuals are not myopic in the model and consider future payoffs with a discount rate β. This is supported by strong evidence of anticipation and intentionality in sexual initiation, found by Kinsman et al. (1998). Consequently, because the model ends with high school graduation but the payoff to virginity status continues, non-trivial 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 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.[16]

I express lifetime utility using the Bellman representation, with value functions denoted $V_a(y_{t-1}, Y_{t-1}, \omega, \epsilon)$. The vector Y_{t-1} (8×1) contains the nonvirginity rates by gender in each of the four grades in high school; this is the aggregate state of the local market. The arguments of the value functions, along with the individual's gender and age, constitute the information set. The value function for a nonvirgin ($y_{i,t-1} = 1$) has an analytical expression as $V_{a_{it}}(1, Y_{t-1}, \omega_i, \epsilon_{it}) =$

$$u_{it} + \omega_i + \epsilon_{it} + \sum_{s=1}^{A-a_{it}} \beta^s [\mathrm{E}_t u_{i,t+s} + \omega_i] + \beta^{(A-a_{it}+1)} \frac{\alpha A + \omega_i}{1 - \beta}, \tag{3}$$

where u_{it} is defined in (2) and E_t denotes the individual's expectation given his or her information set.[17] For a virgin, the value function is a more complicated object that incorporates

[16]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.

[17]To simplify the indexes, note that $a_{it} = t$ for the reference cohort.

the search decision and the arrival rate. It is expressed as $V_{a_{it}}(0, Y_{t-1}, \omega_i, \epsilon_{it}) =$

$$\max_{d_{it}} d_{it} \; \mathrm{E}_t \Big[\lambda_{it} \cdot \Big(u_{it} + \omega_i + \epsilon_{it} + \beta V_{a_{it}+1}(1, Y_t, \omega_i, \epsilon_{i,t+1})\Big) + (1 - \lambda_{it}) \cdot \beta V_{a_{it}+1}(0, Y_t, \omega_i, \epsilon_{i,t+1})\Big]$$
$$+ (1 - d_{it}) \; \beta \mathrm{E}_t V_{a_{it}+1}(0, Y_t, \omega_i, \epsilon_{i,t+1}). \tag{4}$$

The first line 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 second line is 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 $(Y_{it}, Y_{i,t+1}, \dots)$ and arrival rates $(\lambda_{it}, \lambda_{i,t+1}, \dots)$. 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 expected arrival rates can be derived based on the decision rule for the opposite gender. The search decisions among the opposite gender depend on their state variables $(a_{jt}, y_{j,t-1}, Y_{t-1}, \omega_j, \epsilon_{jt})$. Given Y_{t-1}, it is possible to integrate the decision rule over the distributions of ω_j and ϵ_{jt}, along with the various possible assigments of individual virginity statuses $y_{j,t-1}$ that would correspond to the group nonvirginity rates in Y_{t-1}. This yields a distribution of N_{it}, the share of searchers, which in turn gives the distribution of λ_{it} based on (1). How I implement this is explained in sections 3.2 and A.1.

For the beliefs about the evolution of Y_t, I use an approximation to fully rational beliefs that is similar to 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 following specification:

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

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 6). 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.[18] 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.[19]

Finally, the expected costs and benefits of sexual activity embodied in the permanent individual preference term ω_i may relate to the probability of initiation prior to the ninth grade. Because these initiation 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.[20] 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 peer group) indicates a higher ω_i for the individual. Second, if ω is uncorrelated but there are common opportunities to initiate sex prior to the ninth grade, the distribution of ω among the remaining virgins is affected by selection.

[18]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.

[19]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.

[20]Of course, the vector Y_0 contains all grades because they are needed as part of the state space.

In addition to Y_0, the conditional distribution of ω also depends on a vector of exogenous, permanent individual characteristics, x. This is for the empirical implementation, to incorporate observable attributes that relate to the expected costs and benefits of sex. I specify ω to have finite support, $\omega \in \{\omega^k\}_{k=1}^{\kappa}$, which assumes there are κ "types" when it comes to sexual initiation. The conditional distribution of ω, for virgins at the beginning of ninth grade, is specified as a multinomial logit:

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

where the parameters for the first type are normalized to zero.

3.2 Solving the Model

Given beliefs about the evolution of Y_t, the individual decision problem solves much like a single-agent dynamic problem. This simplification occurs because the current period λ_{it} drops out from the decision rule, so there is no simultaneous game to be solved each period. To show this result, I rearrange (4) to

$$\begin{aligned} \max_{d_{it}} \ & d_{it} \cdot \mathrm{E}_t\lambda_{it} \cdot \Big(u_{it} + \omega_i + \epsilon_{it} + \beta \mathrm{E}_t V_{t+1}(1, Y_t, \omega_i, \epsilon_{i,t+1}) - \beta \mathrm{E}_t V_{t+1}(0, Y_t, \omega_i, \epsilon_{i,t+1})\Big) \\ & + \beta \mathrm{E}_t V_{t+1}(0, Y_t, \omega_i, \epsilon_{i,t+1}). \end{aligned} \tag{7}$$

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

$$d_{it} = 1 \text{ iff } u_{it} + \omega_i + \epsilon_{it} + \beta \mathrm{E}_t V_{t+1}(1, Y_t, \omega_i, \epsilon_{i,t+1}) > \beta \mathrm{E}_t V_{t+1}(0, Y_t, \omega_i, \epsilon_{i,t+1}). \tag{8}$$

Thus a virgin will search if and only if the value of becoming sexually active exceeds the value of remaining a virgin.[21] This is a standard result in a model with no search cost.

The age-specific value functions for virgins, given by (7), do not have analytical expressions, but they can be numerically constructed by backward recursion starting from the final period which has known terminal values. 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 in order to account for the search decisions of opposite-gender virgins, which are embedded in the arrival rate (λ_{it}). An exact calculation for the expected arrival rate would use the decision rule in (8), and integrate over the unobserved values of ω_j and ϵ_{jt}. 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 in the model depends on the individual characteristics that affect the distribution of ω and on 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 in 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

[21]It is interesting to note that the criterion would be the same in a decision about accepting an offer to have sex. However the reason to model the decision as search rather than offer acceptance is that search behavior produces an endogenous supply of partners.

(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 re-estimating 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 6 and table 8). 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.[22] This avoids the need to integrate over a distribution in each future period 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 paremeters. This greatly reduces the computational burden of estimation.

[22]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.

3.3 The need for a dynamic model

The model shows why only cross-sectional variation in gender-specific nonvirginity rates cannot separately identify the effects of peer norms and partner availability. Consider 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 peer nonvirginity would also indirectly increase the supply of partners in equilibrium. This is clear in the dynamic context: in period 1, the nonvirginity rate among peers exogenously increases, and the search behavior of these 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; and finally 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 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 1994-95 school year, when the first wave was conducted. The second round of interviews (Wave II) followed up with respondents one year later in April through August of 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 selfdirected 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.[23] Those who say yes are then asked to report the month 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 month 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 1994-95 or 1995-96 school years, who were selected for in-home interviews.[24] Add Health contains 17,657 such individuals, who are in grades 8 through 12 during the first round of interviews in 1994-95. I use the grade in that academic year to refer to separate "cohorts." I exclude 2,635 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 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.

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 one or two 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 cohort.[25] These graphs show that a large portion of individuals initiate sex during high

[23] 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, page 1.)

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

[25] 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.

school. The share of nonvirgins among boys increases from just over 26 percent at the beginning of ninth grade to just under 64 percent at the end of twelfth grade, and among girls it increases from 20 percent to 62 percent. Thus about 40 percent of the population initiates sex during the four years of high school.

Data on the characteristics (x) that affect the distribution of the individual preference term (ω) come from Wave I. I use black race, parental education, and sibling status, because they are predetermined and have been shown to predict early sexual initiation in other work.[26] 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. These variables have relationships with the expected costs and benefits of sexual activity, for example through expected labor market outcomes or through information from older siblings. Table 1 gives the unweighted and weighted means of these indicators (i.e., the sample shares). There are additional variables which I use in the IV estimation described below, which are not included in the model due to the computational burden of additional parameters. These are indicators for Hispanic ethnicity, mother currently married, a foreign-born parent, high household income (above $50,000), and early menarche for girls (before age 12). The weights make little difference except for the shares with black race, Hispanic ethnicity, or a foreign-born parent, which reflect oversamples in the sample design.

Table 2 shows the raw correlation between individual virginity status and the nonvirginity rates for each gender and grade at the same school, assessed in the last observation period (the spring quarter of 1996).[27] The bolded numbers along the diagonal give the correlations within peer groups, which 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). This provides support for the definition of peer groups

[26]See section 2. Also, Widmer (1997) and Argys et al. (2006) give evidence that individuals with older siblings tend to initiate sex earlier.

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

by grade. Overall, there are large correlations in virginity status within schools, about 0.2 in magnitude, which indicates substantial variation in nonvirginity rates across schools.

The bolded 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; and for girls, they are boys in the same grade and the next two older grades.[28] 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, and to create variation within schools in the shares of nonvirgins and searchers on the supply side of the market. Partners from outside these groups, 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 between individuals and their supply groups as used for estimation are larger than the correlations with the excluded grades of the opposite gender.

4.1 Evidence of the composite effect

Before estimating the search and matching model, I briefly present estimates of the composite effect of social interactions based on standard methods.[29] This demonstrates the presence of social interactions in my data using very different identifying assumptions than in the structural estimation. I estimate regressions for virginity status at the spring quarter of each grade in high school, which have the following form:

$$y_{imat} = \pi_{0at} + \pi_{1a}' x_i + \pi_{2a} \bar{y}_{(-i)mt} + w_{ma} + e_{imat}, \tag{9}$$

[28]The supply groups do not need to be symmetric in the model or in reality 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.

[29] A more extended analysis is presented in an earlier version of this paper. See section 5 of the December 2010 version, available at: `http://www.andrew.cmu.edu/user/sethrs/SexInitiation_Dec2010.pdf`.

where m indexes schools, a is grade, and t is calendar time.[30] The variable $\bar{y}_{(-i)mt}$ contains the nonvirginity rate in the reference group, defined below, and w_{ma} is a school-by-grade fixed effect. Estimation is via 2SLS. As is standard, I use means of the individual characteristics within the reference group ($\bar{x}_{(-i)m}$) to instrument for $\bar{y}_{(-i)mt}$ (e.g., Gaviria and Raphael 2001). Identification relies on the school-by-grade fixed effects to eliminate any violations of the exclusion restrictions. This assumes that *differences* in mean peer characteristics from one cohort to the next are exogenous, a strategy introduced by Hoxby (2000) and used by Hanushek et al. (2003) and Lavy and Schlosser (2011), among others.

These regressions cannot be interpreted as approximations of the search and matching model. As explained in section 3.3, a static model cannot separate the effects of peer norms and partner availability. Accordingly, because it would be hard to interpret any differences between a within-gender effect and a cross-gender effect, these regressions include only a single reference group. However to explore different sources of variation, I consider two alternative constructions: one is the peer group as defined in the model, i.e., by gender and grade; the other pools the peer group with the opposite-gender supply groups described earlier.[31] Also, I use contemporaneous peer outcomes in the regressions rather than lagged outcomes, because this is the standard formulation in the literature.[32]

Table 3 presents the estimates of π_2 by gender and grade, using the alternative reference groups described above.[33] This is the effect of the nonvirginity rate in the reference group on the probability that an individual is sexually experienced—what I call the composite effect of social interactions. For example, in columns 3 and 7 of panel A, the point estimates imply that a 10 percentage-point increase in the nonvirginity rate among (same-gender) peers

[30] Note that although individual outcomes are subscripted with t, an individual appears only once in the sample for each regression. The term π_{0at} allows for differences in the average outcomes of each cohort.

[31] The regression coefficient for either construction combines the demand and supply effects contained in the model, but the exact combination will be different under the two constructions.

[32] Regressions using lagged outcomes produce similar results. See note 29 for the reference.

[33] Full results for each of these specifications are contained in appendix tables A-1 and A-2.

raises the probability of being sexually experienced by 7.4 percentage points for boys and 9.6 points for girls in the eleventh grade. Overall these results provide strong evidence that school-based social interactions have an effect on sexual initiation for both boys and girls.

5 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 per-period probability of the transition from virginity to nonvirginity. Because search decisions are unobserved, this transition probability is given by the product of the probabilities of searching and of finding a partner. In order to separate the arrival rate from the search probability without relying on functional form, I take advantage of additional data on the arrival of subsequent partners after the first. Thus the likelihood function includes individual contributions for both the duration to first sex and the arrival of subsequent partners for nonvirgins.

5.1 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, its CDF denoted Φ, this product is

$$\begin{aligned} L_{it}(\omega) \equiv \; & \Phi\left[u_{it} + \omega + \beta \mathrm{E}_t V_{a_{it}+1}(1, Y_t, \omega, \epsilon_{i,t+1}) - \beta \mathrm{E}_t V_{a_{it}+1}(0, Y_t, \omega, \epsilon_{i,t+1})\right] \\ & \cdot \int \lambda_{a_{it}}(N_{it}) \; f(N_{it}) \; \mathrm{d}N_{it}. \end{aligned} \tag{10}$$

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

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), in order 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]

$$L_i(\omega) \equiv L_{it_i^*}(\omega) \cdot \prod_{s=1}^{t_i^*-1} [1 - L_{is}(\omega)]. \tag{11}$$

Adding across types, the probability given for the duration is

$$\mathcal{L}_i \equiv \sum_{k=1}^{\kappa} \pi_{k|Y_{m0},x_i} L_i(\omega^k) \tag{12}$$

where $\pi_{k|Y_{m0},x_i}$ is defined in (6) and m indexes schools.

In addition to the likelihood contributions for the durations to first sex, the likelihood function contains individual contributions from nonvirgins in order to improve the estimation of the arrival rate parameters λ_0 and λ_1. This draws on data from the sexual histories reporting 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 subsequent sexual partners after the first one directly identifies the raw arrival rate. To limit departures from 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

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

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

$$\mathcal{A}_i \equiv \mathrm{E}[\lambda_{a_{it_i^{**}}}(N_{it_i^{**}})] \prod_{s=t_i^*}^{t_i^{**}-1} (1 - \mathrm{E}[\lambda_{a_{is}}(N_{is})]) \tag{13}$$

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 (10).[36] I restrict to individuals with $y_{i0} = 0$ (initial virgins) in order 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 (13) simply multiply with the likelihood contributions in (12). 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 who 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 (Gouriéroux and Monfort 1996). The variance

[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.

approximation is calculated via numerical differentiation.[37]

5.2 Identification

This analysis of identification has two parts. First I show the identification of the parameters of the model, and then I discuss more generally how my strategy addresses identification problems that typically arise with social interactions.

5.2.1 Identification of the structural parameters

The observed data contain 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 x 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 $Y_{kt} = n^{-1} \sum y_{kit}$ 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 $\mathbf{x}_s = \{x_j :$

[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.

$j \in S_a(i)\}$ 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]

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

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

where Λ is the logistic function. The random variable N_t is the fraction of searchers

[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.

[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.

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]

5.2.2 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 group means of individual characteristics have a direct effect on individual behavior, because they also determine mean group behavior. Brock and Durlauf (2001a, 2001b) show that this does not apply in binary choice models and certain other nonlinear models, because mean group outcomes are no longer linearly dependent on mean characteristics. Although my model is nonlinear, the use of an approximation to rational expectations may raise the question of whether there is a reflection problem. However the approximation ψ is in fact nonlinear, 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).

Moreover, it is worth noting that the reflection problem would not apply in my case even if there were a linear dependence. The reason is that the standard assumption of stable preferences over time provides an identifying restriction on the key parameters. In Manski's (1993) model and related models considered in Brock and Durlauf (2001b), the group-level variables are essentially arbitrary, so there is no theory to inform their direct effects on behavior compared with the effect of the expected group outcome. In my model, the group-

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

level variable (Y_{t-1}) is the lag of the expected group outcome ($\mathrm{E}_t[Y_t]$). The effects of the lag and the expected outcome on utility, and hence on behavior, are the same up to the (known) discount factor. (See appendix B for a more formal discussion.)

Turning to the other problems, the definition of peer groups by grade is intended to avoid selection bias, which would occur if endogenous friendships or activities like sports teams were used. This assumes that individuals do not systematically skip or repeat grades in order to affect their chances of sexual initiation. Hoxby (2000), Hanushek et al. (2003), and other authors similarly assume that school cohorts are exogenous. The more general problem of correlated unobservables among peers is addressed by making the distribution of ω a function of initial nonvirginity rates, Y_0. This controls for correlated omitted variables that are time invariant, under the assumption that Y_0 is a sufficient statistic for their effects.

With this strategy, the variation used to identify a peer influence on search decisions comes from differences in Y_{it} across peer groups conditional on initial nonvirginity rates, Y_0. Two groups with the same Y_0 would have the same distribution of ω for their members (controlling for individual characteristics), but different values of Y_{it} in later periods would produce different search probabilities due to the peer effect. Thus, intuitively, the estimate of γ is based on the magnification of small differences over time given similar initial conditions.

6 Results and Counterfactual Simulations

The estimated model fits the observed patterns in sexual initiation, and it finds meaningful differences between the two mechanisms of peer norms and partner availability. Figure 2 compares the observed growth of nonvirginity rates for a synthetic cohort ("8-11 Observed") against predictions from the model. The predicted line ("9 Predicted") shows the ninth grade cohort projected through the end of twelfth grade. This prediction is formed by starting with observed virginity status in the initial time period (the summer before the ninth grade cohort entered high school, 1994Q3), and then simulating the outcomes for all cohorts going forward. Thus by the time the ninth grade cohort reaches the end of high school, the prediction is fifteen 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 tenth grade, so their predicted outcomes in eleventh and twelfth grades are shown against the observed outcomes of older cohorts.

The structural parameters 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. The age parameter, α, is twice as large for girls compared with boys, which suggests that girls are more influenced by individual development. 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 is one parameter for each of the three grades that provide the endogenous supply of partners for an individual, as explained in section 4. These three parameters are jointly significant for boys but not for girls. Estimating with the alternative matching technology yields similar values for γ but shows λ_1 essentially equal to zero (table A-3). The difference between these results could be interpreted as evidence that individuals meet each other randomly at school, regardless of their search decisions. This would make the arrival rate a function of the proportion of searchers—i.e., those who are willing to match—among the opposite gender (as in the the main specification). It also suggests that any constraint on the number of partners per period is not binding, at least on the margin used to estimate λ_1. 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.

Two types are sufficient to fit the observed growth of nonvirginity rates during high school, as well as differences in these growth trends along various observable characteristics (not shown).[42] I refer to these as "low" and "high" types, where $\omega^L < \omega^H$. While table 4 contains the parameters, table 5 shows the probability of being high type for the specified group (virgins at the beginning of ninth grade), along with the average partial effects of the individual characteristics that affect this probability.[43] Roughly half are high type, with a higher rate for boys than girls. The partial effects are qualitatively similar to the coefficients estimated for these variables in the 2SLS exercise (see tables A-1 and A-2), including the fact that black race has a large effect for boys but not girls.

To help interpret the key parameters 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. 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 or 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 one-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.[44] This is 17% (14%) of the average search probability among boys (girls). Table 7 shows that the arrival rate of partners is similar for boys and girls and is fairly constant across grades.[45] The average arrival rate is about 0.1, which corresponds to an expected wait of 10 periods or 2.5 years to find a partner. The marginal effects are smaller

[42] Arcidiacono, Khwaja, and Ouyang (2011) also find that two types are sufficient in their work on adolescent sexual behavior.

[43] 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.

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

[45] The arrival rate does not automatically increase over time, even though nonvirginity rates are monotonically increasing, because the constant term λ_0 varies by grade. In addition, not all grades have three supply groups available at school; for example, there are no older groups for twelfth graders.

in magnitude than those in table 6, but the combined effect of search behavior among the supply groups can be large for boys. For example, an increase of one standard deviation in the search behavior within each of the three supply groups raises the arrival rate of partners for a tenth grade boy by 0.02, or 18% of the average rate. For girls, the underlying parameters are not jointly significant so any apparent marginal effects may only reflect sampling noise.

Finally, table 8 gives the parameters of the approximation of equilibrium beliefs (equation 5) that was used to estimate the model (columns labeled "observed"). After estimation of the structural parameters, I then re-estimate these regressions with data simulated from the model to check that the approximation is consistent with the model (columns labeled "simulated"). The new coefficients are generally quite close to the ones in the 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 naıve correlation between the observed and simulated data, but not any further variance or covariance 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.

6.1 Counterfactual Simulations

The counterfactuals in figures 3 through 5 demonstrate the impacts of peer norms and partner availability and simulate the effects of two interventions. Each figure has two graphs, for boys and 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 some modification to produce the simulation ("9 Simulated"). For reference, the gray line in these graphs ("8-11 Observed") shows the observed nonvirginity rates for the synthetic cohort, but the relevant comparisons are between the estimated model and the simulations. In each counterfactual, equilibrium beliefs must be revised in order 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 in the beliefs approximation converge, which usually occurs in fewer than 10 iterations.

The counterfactuals in figures 3 and 4 demonstrate the overall impact of school-based peer norms and partner availability on sexual initiation during high school, by showing what happens if their effects were eliminated. Figure 3 removes the influence of peer norms by setting $\gamma = 0$. The result is that the number of boys (girls) who become sexually active is reduced by 10 (9) percentage points, which is 26% (20%) of the total during high school. The relative impact 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% for girls. Figure 4 eliminates the effect of opposite-gender search behavior on the arrival rate by setting the vector $\lambda_1 = 0$. This shows 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 share of boys who become sexually active during high school falls by 0.14 (37% of the total). This represents an upper bound on the potential impact of single-sex schools. To the extent that boys would compensate by increasing their search effort in the external market, a choice which is outside the model, the impact would be reduced.

Figures 5 simulates isolating the ninth grade from the rest of high school. This is accomplished by zeroing out the parameters that apply to older supply groups in the arrival rate for ninth graders (i.e., $\lambda_{g1,2}$, $\lambda_{g1,3}$, and $\lambda_{b1,3}$) and the parameter for the younger supply group for tenth grade boys ($\lambda_{b1,2}$). This decreases sexual initiation in the ninth grade by about 14% for both boys and girls, although the impact on girls is based on poorly measured parameters. The effect dissipates rapidly for girls, but it persists somewhat for boys due to the additional supply reduction for tenth-grade boys.

7 Conclusion

This paper contributes to the literature on social interactions by estimating a model that formally distinguishes between two potential mechanisms in an application on sexual initiation. I estimate an equilibrium search and matching model for first sexual partners, which contains an effect of peer-group norms on the demand for sex and captures the effect of opposite-gender search behavior on the arrival rate. The results indicate that peer norms have a substantial effect on the timing of sexual initiation for both boys and girls. Without an effect of peer norms, the number of individuals who become sexually active during high school falls by 26% for boys and 20% for girls. The availability of partners at school also has a large effect on boys, although not on girls.

The estimation of a dynamic equilibrium model is necessary in order to disentangle the effects of the two mechanisms. As I show, exogenous changes in demand and supply cannot do this in a static context because any change on one side of the market indirectly shifts the supply curve from the other side, due to the presence of preference interactions on each side. In addition, the estimation of structural parameters allows for counterfactual simulations that are, in theory, robust to changes in the policy environment. I consider two types of interventions, which correspond to the two mechanisms: educational programs to reduce the influence of peer norms, and the isolation of the ninth grade from older grades. The large effect of peer norms indicates the potential for the former to have an impact, while a simulation of the latter shows a relatively small reduction in sexual initiation.

This approach to investigating mechanisms behind social interactions, of course, has limitations as well as advantages. Certain functional form assumptions and approximations are required in order to estimate the structural parameters of the model. However I confirm the presence of social interactions using a standard method, which has different identifying assumptions than in the structural estimation. Among the results from that exercise, it is interesting to note that when the reference group includes the opposite gender as well as the peer group, there is a larger composite effect for boys than for girls (table 3). Although it is only suggestive, this corresponds to the structural estimates which find an effect of opposite gender search behavior on the arrival rate for boys but not for girls.

Future work on this topic would benefit from data over time on reported intentions to initiate sex.[46] If reliable, they could be treated as observed search decisions, which would remove the need to rely on subsequent partners to identify the arrival rate separately from the search probability. Observing search decisions would also make it easier to include heterogeneity in the arrival rate, and to allow for possible competition for partners between individuals of the same gender.

A Approximations

A.1 Arrival Rate for Model Solution

As noted in section 3.2, it would be difficult to make an exact calculation for the expected arrival rate at randomly selected solution points, in order to construct the value functions in (4). Instead, I approximate the decision rules among the opposite gender and use this approximation to simulate an expected value for the arrival rate.

This procedure uses auxiliary regressions that relate the probability of search among both virgins and nonvirgins in a group to the lagged nonvirginity rate for that group. These regressions are made separately for each gender and quarterly "age" in high school. To construct the regression coefficients, I start with initial values which allow the model to be solved and thus generate search probabilties for the entire sample, and then iterate with regressions of these search probabilities on the observed lagged nonvirginity rates. This approach should well approximate the information structure specified in the model, because the state space does not include any information about the opposite gender apart from their lagged nonvirginity rates in Y_{t-1}.

To calculate the expected arrival rate ($\mathrm{E}_t\lambda_{it}$) to go into (4) or (7) with this approach, I first

[46]Add Health contains questions on the components of an ideal romantic relationship, which indicate whether the individual wants to become sexually active. However, these are assessed only at the time of the interviews, so they would not allow the identification strategies used in the estimation of the model.

use the appropriate age-specific regression to assign a probability of search (not conditional on virginity status) to each person in the supply groups, based on the value of $Y_{k,t-1}$ for their group (k). Then I use a series of uniform draws to simulate their behavior several times, which gives a number of realizations for N_{it}. Finally I average over the resulting values of $\lambda(N_{it})$ to calculate an expected value for λ_{it}. The remainder of the solution algorithm for the individual problem is standard.

A.2 Arrival Rate for Estimation

I use Monte Carlo integration to approximate the expected arrival rate in (10), because the search decisions of virgins in the supply groups $S(i)$ are unobserved and depend on their individual shocks. The procedure is as follows. For each simulation round, $r \in 1 \dots R$, simulate a search decision, d^r_{jt}, for each virgin, j, in the three supply groups (to get $\{d^r_{jt} : j \in S(i),\ y_{j,t-1} = 0\}$). This proceeds by drawing ω^r_j from the appropriate distributions and then comparing the type-specific search probability for individual j, given by $\Phi[\dots]$ in (10), against a pseudorandom uniform draw. Combining these simulated search decisions of virgins with the known search behavior of nonvirgins (i.e., they all search) yields N^r_{it}. Then averaging $\lambda_{a_{it}}(N^r_{it})$ across simulation rounds produces an approximation for the expected arrival rate.

A.3 Type Distribution for Virgins First Observed after the Ninth Grade

As noted in section 5.1, I need to update the distribution of ω for cohorts that are first observed after ninth grade. To do this I create approximate, type-specific hazard rates, which combine to give the probability, for each type, of still being a virgin when the individuals are first observed. Then I can use Bayes Rule to update the initial distribution of ω (for ninth graders) to the appropriate grade.

The approximate, type-specific hazard rates are created with data from the younger cohorts at the individuals' schools, which relies on a steady state from one cohort to the next.[47] I regress the type-specific transition probabilties ($L_{it}(\omega)$) of the younger cohorts on their relevant state variables, which include the lagged nonvirginity rates by gender and grade ($Y_{m,t-1}$). I then use these regressions to predict type-specific hazard rates for the older

[47]The steady state assumption appears elsewhere, notably in the use of an aggregate law of motion estimated from current data to function as beliefs about the future.

cohorts before the observation period (i.e., when they were in earlier grades in high school). In these predictions, the current nonvirginity rates among younger cohorts substitute for the unobserved rates among older cohorts in previous years. The predicted hazard rates yield the probability of remaining a virgin for each type, and then I use Bayes Rule to update the initial distribution of ω for each individual in the older cohorts.

The exact procedure is:

i) Regress $L_{it}(\omega)$ on $Y_{m,t-1}$ and $\bar{x}_{s(i)}$, with separate approximations for each gender and age (i.e., quarter within grade).

ii) Project Y_{m0} forward using the approximation ψ, to create a sequence as long as the unobserved time span. For example, for someone first observed at the beginning of eleventh grade, Y_{m0} would be projected for two years (eight periods).

iii) Predict $\widehat{L}_{it}(\omega)$ for the unobserved periods using the regressions from step (i), and the generated sequence of $Y_{mt}, t < 1$, from step (ii).

iv) Define the approximate, type-specific probabilities of still being a virgin in the initial observation period as

$$\widehat{P}_i^0(\omega) \equiv \prod_{t=1-a_{i0}}^{0} [1 - \widehat{L}_{it}(\omega)].$$

v) Finally, update the individual's type distribution with

$$\Pr(\omega_i = \omega^k \mid Y_{m0}, x_i, a_{i0}, y_{i0} = 0) = \frac{\widehat{P}_i^0(\omega^k) \cdot \pi_{k|Y_{m0},x_i}}{\sum_{l=1}^{\kappa} \widehat{P}_i^0(\omega^l) \cdot \pi_{l|Y_{m0},x_i}}.$$

The circularity in this procedure is resolved by starting with some initial guess for the regressions that produce $\widehat{L}$ and then iterating. In practice, these approximations converge very quickly (within three iterations).

B Stable Preferences and the Reflection Problem

Here I show how the assumption of stable preferences over time provides an identifying restriction that can address the reflection problem. First we need to modify the model so that there is a linear dependence between the effects of Y_{t-1} and $\psi(Y_{t-1})$ on behavior. Accordingly, suppose ψ is a linear function ($\psi(Y_t) = \psi Y_t$), and consider a once-and-for-all decision to initiate sex in period 2 with the arrival rate set equal to one. In addition, let Y_t be a scalar and simplify the heterogeneity as $\omega_i = a'x_i + bY_0$, so that flow utility for a nonvirgin is $a'x_i + bY_0 + cY_{t-1} + e_{it}$.[48] Then the expected payoff to searching in period 2 is the discounted sum

$$\sum_{t=2}^{\infty} \beta^{t-2} \left(a'x_i + bY_0 + c\mathrm{E}_2[Y_{t-1}]\right)$$

(plus e_{it}), where $\mathrm{E}_2[Y_{t-1}]$ denotes the fully rational expectation for Y_{t-1} given the information set at period 2 (which includes Y_1). The approximation replaces $\mathrm{E}_2[Y_{t-1}]$ with $\psi^{t-2}Y_1$, and so the expression above becomes $(1-\beta)^{-1}(a'x_i + bY_0) + (1-\beta\psi)^{-1}cY_1$. With ψ identified from the distribution of (Y_1, Y_2), and an assumed value for β, the parameters a, b, and c are identified.

The difference with Manski's (1993) result is that the relationship between the effects of Y_1 and $\psi^{t-2}Y_1$ on behavior are known. The latter equals the former scaled by powers of β under the assumption of stable preferences. By contrast in Manski's static model, and the models considered in Brock and Durlauf (2001b) and Blume et al. (2011), the common group variable is essentially arbitrary, call it X_g. Because there is no theory to inform the relationship between the effects of X_g and $\mathrm{E}[Y|X_g]$ on behavior, there is another parameter to estimate which creates the identification problem.

References

[1] Abma, J. C., G. M. Martinez, W. D. Mosher, and B. S. Dawson. 2004. "Teenagers in the United States: Sexual Activity, Contraceptive use, and Childbearing, 2002." *Vital and Health Statistics*, 23(24).

[2] Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games." *Econometrica*, 75(1): 1-53.

[3] Arcidiacono, Peter, Ahmed Khwaja, and Lijing Ouyang. 2011. "Habit Persistence and Teen Sex: Could Increased Access to Contraception have Unintended Consequences for Teen Pregnancies?" *Journal of Business and Economic Statistics*, forthcoming.

[4] Arcidiacono, Peter, G. Foster, N. Goodpaster, and J. Kinsler. 2011. "Estimating Spillovers using Panel Data, with an Application to the Classroom." Working paper.

[5] Arcidiacono, Peter, Andrew Beauchamp, and Marjorie McElroy. 2011. "Terms of Endearment: An Equilibrium Model of Sex and Matching." Working paper.

[48] Blume et al. (2011) consider similar payoffs in a dynamic linear model (pp. 869-870).

[6] Argys, Laura M., Daniel I. Rees, Susan L. Averett, and Benjama Witoonchart. 2006. "Birth Order and Risky Adolescent Behavior." *Economic Inquiry*, 44(2): 215-233.

[7] Bajari, Patrick, C. L. Benkard, and Jonathan Levin. 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, 75(5): 1331-1370.

[8] Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. 2011 "Identification of Social Interactions." In *Handbook of Social Economics*, Volume 1, ed. Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 853-964.

[9] Brock, William A. and Steven N. Durlauf. 2001a. "Discrete Choice with Social Interactions." *Review of Economic Studies*, 68(2): 235-260.

[10] ——. 2001b. "Interactions-Based Models." In *Handbook of Econometrics*, Volume 5, ed. J. J. Heckman and E. Leamer, 3297-3380.

[11] Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2011. "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects." Working paper.

[12] Case, Anne C. and Lawrence F. Katz. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." National Bureau of Economic Research Working Paper Series, No. 3705.

[13] Centers for Disease Control and Prevention (CDC). 2008a. "Youth Risk Behavior Surveillance: United States, 2007." *Morbidity and Mortality Weekly Report*, 57(No. SS-4).

[14] ——. 2008b. *Sexually Transmitted Disease Surveillance, 2007.* Atlanta, GA: U.S. Department of Health and Human Services.

[15] Chesson, Harrell W., John M. Blandford, Thomas L. Gift, Guoyu Tao, and Kathleen L. Irwin. 2004. "The Estimated Direct Medical Cost of Sexually Transmitted Diseases among American Youth, 2000." *Perspectives on Sexual and Reproductive Health*, 36(1): 11-19.

[16] Choo, Eugene and Aloysius Siow. 2006. "Who Marries Whom and Why." *Journal of Political Economy*, 114(1), 175-201.

[17] Clark, Andrew E. and Youenn Lohéac. 2007. " 'It wasn't me, it was them!' Social Influence in Risky Behavior by Adolescents." *Journal of Health Economics*, 26(4): 763-784.

[18] Cooley, Jane. 2010. "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box." Working paper.

[19] Eckstein, Zvi, and Kenneth I. Wolpin. 1990. "Estimating a Market Equilibrium Search Model from Panel Data on Individuals." *Econometrica*, 58(4): 783-808.

[20] Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy*, 100(5): 966-991.

[21] Fletcher, Jason M. 2007. "Social Multipliers in Sexual Initiation Decisions among U.S. High School Students." *Demography*, 44(2): 373-388.

[22] Gaviria, Alejandro and Steven Raphael. 2001. "School-Based Peer Effects and Juvenile Behavior." *Review of Economics and Statistics*, 83(2): 257-268.

[23] 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.

[24] Hoffman, S. 2006. "By the Numbers: The Public Costs of Teen Childbearing." National Campaign to Prevent Teen Pregnancy.

[25] Hotz, V. J. and Robert A. Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies*, 60(3): 497-529.

[26] Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." National Bureau of Economic Research Working Paper 7867.

[27] Jackson, C. K. and E. Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics*, 1(4): 85-108.

[28] Keane, Michael P. and Kenneth 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.

[29] Keane, Michael P. and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy*, 105(3): 473-522.

[30] 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.

[31] Krusell, Per and Anthony A. Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106(5): 867-896.

[32] Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, 3(2): 1-33.

[33] Lee, Donghoon, and Kenneth I. Wolpin. 2006. "Intersectoral Labor Mobility and the Growth of the Service Sector." *Econometrica*, 74(1): 1-46.

[34] Levine, Phillip B. 2001. "The Sexual Activity and Birth Control Use of American Teenagers." In *Risky Behavior among Youths: An Economic Analysis*, ed. Jonathan Gruber. Chicago and London: University of Chicago Press.

[35] Lucas, Robert. 1976. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conferences on Public Policy volume 1, ed. Karl Brunner and Allan H. Melzer, 19-46. Amsterdam: North-Holland.

[36] Lundborg, Petter. 2006. "Having the Wrong Friends? Peer Effects in Adolescent Substance use." *Journal of Health Economics*, 25(2): 214-233.

[37] Manlove, J., A. Romano-Papillo, and E. Ikramullah. 2004. "Not Yet: Programs to Delay First Sex among Teens." National Campaign to Prevent Teen Pregnancy.

[38] Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531-542.

[39] Markowitz, Sara, Robert Kaestner, and Michael Grossman. 2005. "An Investigation of the Effects of Alcohol Consumption and Alcohol Policies on Youth Risky Sexual Behaviors." *American Economic Review: Papers and Proceedings*, 95(2): 263-266.

[40] Mas, Alexandre and Enrico Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1): 112-145.

[41] Michael, Robert T. and Courtney Bickert. 2001. "Exploring Determinants of Adolescents' Early Sexual Behavior." In *Social Awakening: Adolescent Behavior as Adulthood Approaches*, ed. Robert T. Michael, 137-173. New York: Russell Sage Foundation.

[42] Miller, Brent C., Maria C. Norton, Thom Curtis, E. J. Hill, Paul Schvaneveldt, and Margaret H. Young. 1997. "The Timing of Sexual Intercourse among Adolescents: Family, Peer, and Other Antecedents." *Youth & Society*, 29(1): 54-83.

[43] Moffitt, Robert A. 2001. "Policy Interventions, Low-Level Equilibria, and Social Interactions." In *Social Dynamics*, ed. Steven N. Durlauf and H. Peyton Young, 45-82. Economic Learning and Social Evolution, vol. 4. Washington, D.C.: Brookings Institution Press; Cambridge and London: MIT Press.

[44] Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies*, 61(3): 397-415

[45] Murphy, Kevin M. and Robert H. Topel. 2002. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics*, 20(1): 88-97.

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

[47] Oettinger, Gerald S. 1999. "The Effects of Sex Education on Teen Sexual Activity and Teen Pregnancy." *Journal of Political Economy*, 107(3): 606-644.

[48] Pakes, Ariel, Michael Ostrovsky, and Steven Berry. 2007. "Simple Estimators for the Parameters of Discrete Dynamic Games (with entry/exit Examples)." *RAND Journal of Economics*, 38(2): 373-399.

[49] de Paula, Áureo. 2009. "Inference in a Synchronization Game with Social Interactions." *Journal of Econometrics*, 148(1): 56-71.

[50] Rees, Daniel I., Laura M. Argys, and Susan L. Averett. 2001. "New Evidence on the Relationship between Substance use and Adolescent Sexual Behavior." *Journal of Health Economics*, 20(5): 835-845.

[51] Sabia, Joseph J. 2007. "Reading, Writing, and Sex: The Effect of Losing Virginity on Academic Performance." *Economic Inquiry*, 45(4): 647-670.

[52] Sabia, Joseph J. and Daniel I. Rees. 2008. "The Effect of Adolescent Virginity Status on Psychological Well-being." *Journal of Health Economics*, 27(5): 1368-1381.

[53] Santelli, John S., Javaid Kaiser, Lesley Hirsch, Alice Radosh, Linda Simkin, and Susan Middlestadt. 2004. "Initiation of Sexual Intercourse among Middle School Adolescents: The Influence of Psychosocial Factors." *Journal of Adolescent Health*, 34(3): 200-208.

[54] Sen, Bisakha. 2002. "Does Alcohol-use Increase the Risk of Sexual Intercourse among Adolescents? Evidence from the NLSY97." *Journal of Health Economics*, 21(6): 1085-1093.

[55] Sieving, Renee E., Marla E. Eisenberg, Sandra Pettingell, and Carol Skay. 2006. "Friends' Influence on Adolescents' First Sexual Intercourse." *Perspectives on Sexual and Reproductive Health*, 38(1): 13-19.

[56] Sirakaya, Sibel. 2006. "Recidivism and Social Interactions." *Journal of the American Statistical Association*, 101(475): 863-877.

[57] Trogdon, Justin G., James Nonnemaker, and Joanne Pais. 2008. "Peer Effects in Adolescent Overweight." *Journal of Health Economics*, 27(5): 1388-1399.

[58] Widmer, Eric D. 1997. "Influence of Older Siblings on Initiation of Sexual Intercourse." *Journal of Marriage and Family*, 59(4): 928-938.

Figure 1: Observed Nonvirginity Rates by Quarter within Grade in High School

Figure 1a.

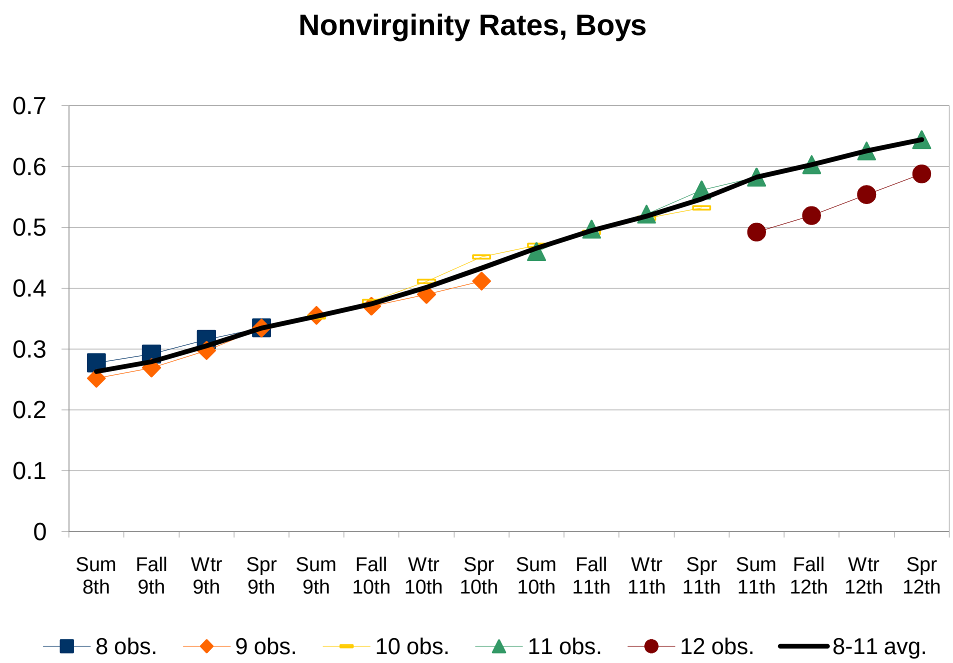
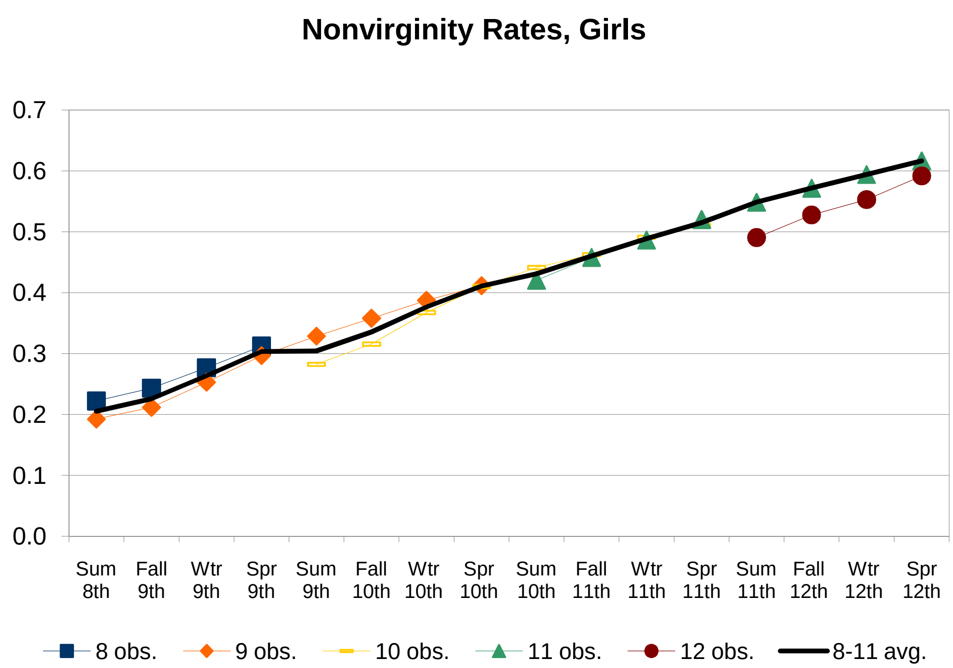


Figure 1b.



NOTES: Each line shows a different cohort, defined by grade in the 1994-95 school year. "8-11 avg." averages across cohorts, to make synthetic cohort.

Figure 2: Model Fit

Figure 2a.

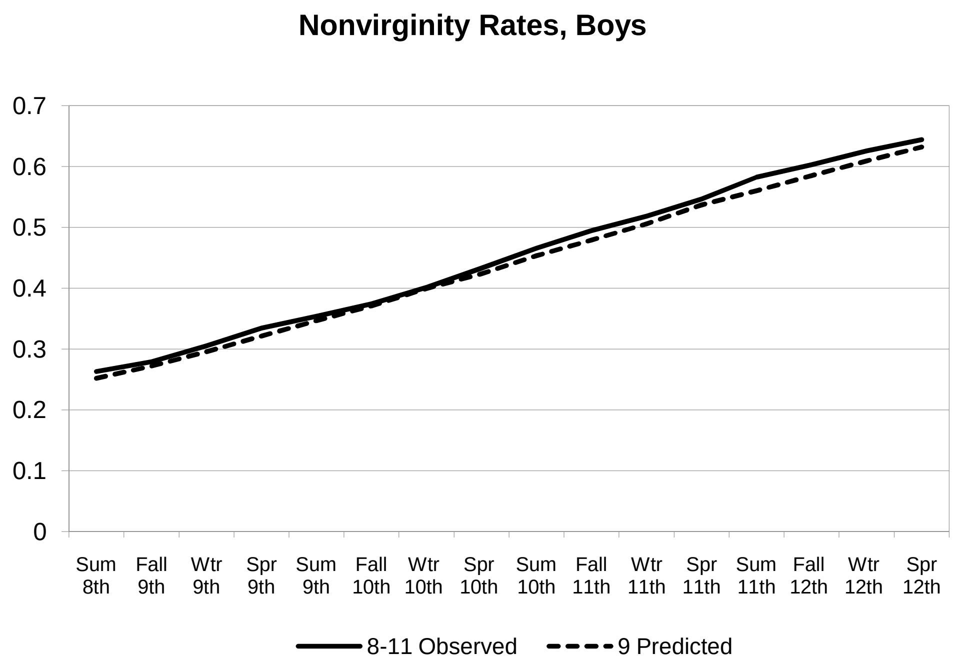
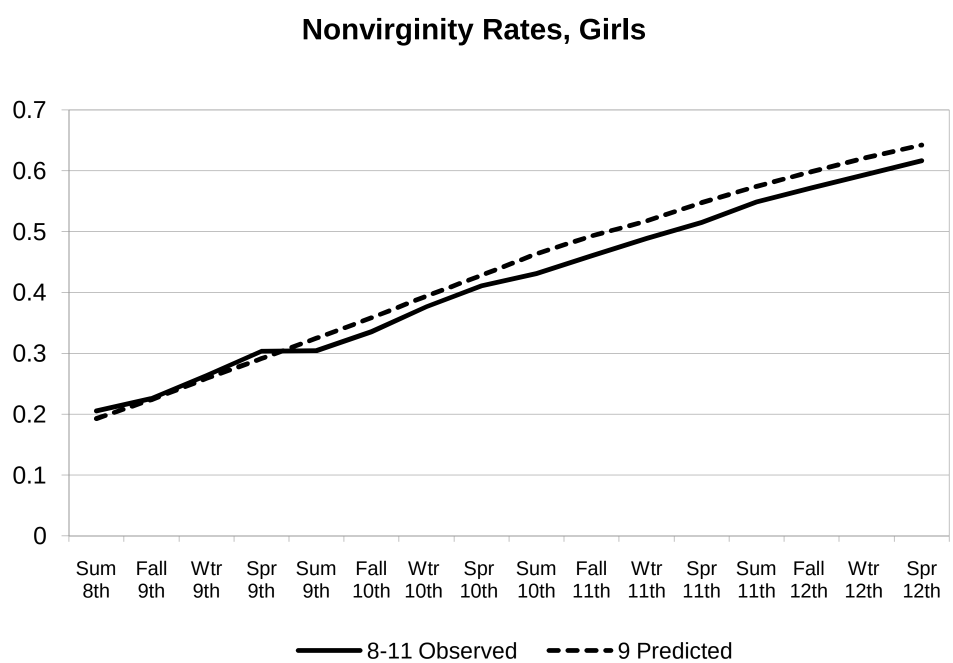


Figure 2b.



NOTES: "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.

Figure 3: Eliminate Effect of Peer Norms on Search Decisions

Figure 3a.

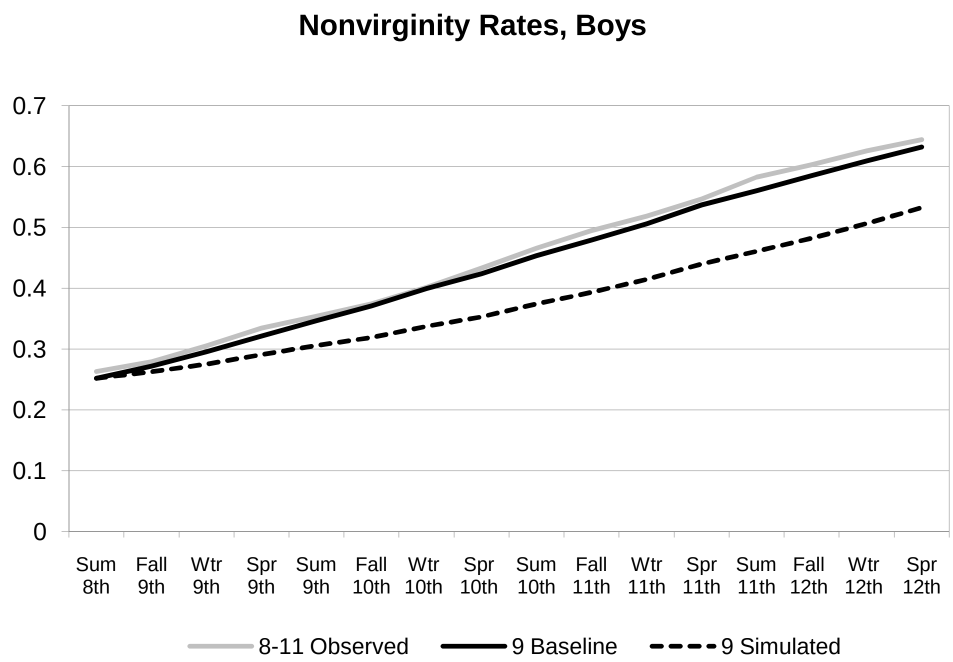
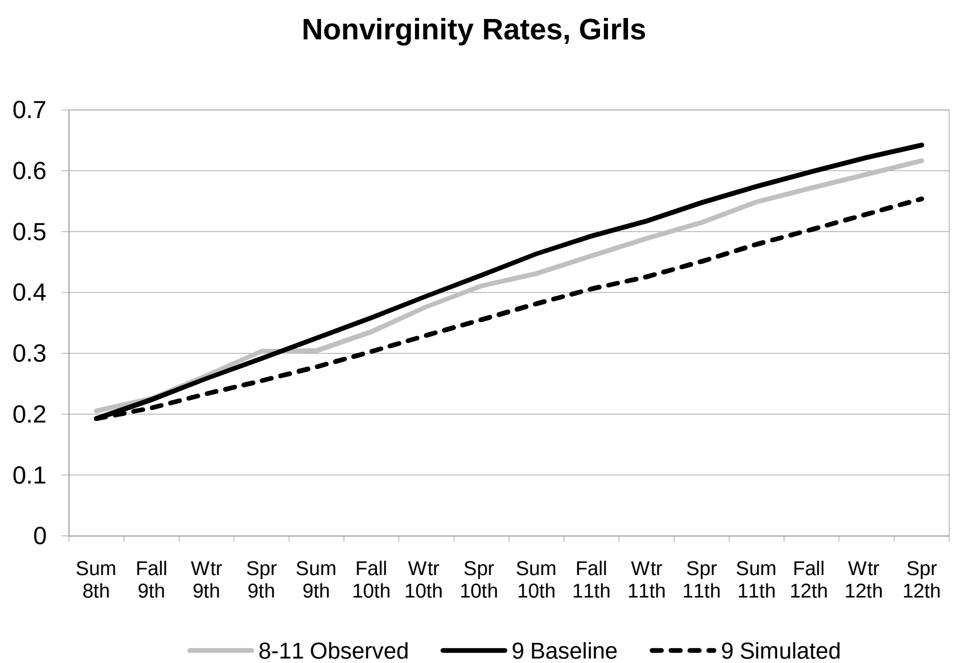


Figure 3b.



NOTES: "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.

Figure 4: Eliminate Effect of Opposite Gender Search Behavior on Arrival Rates

Figure 4a.

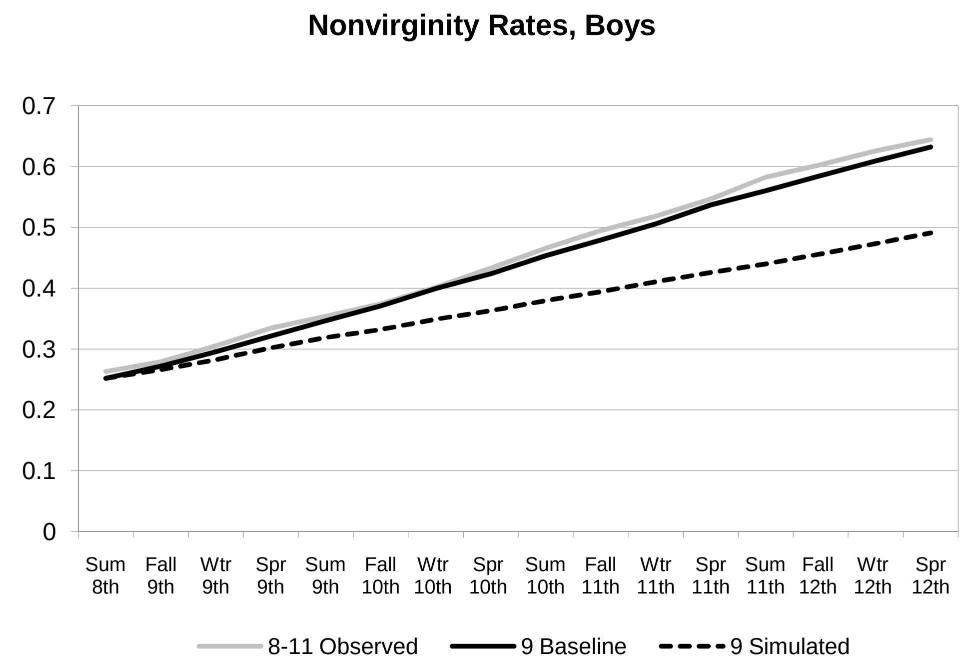
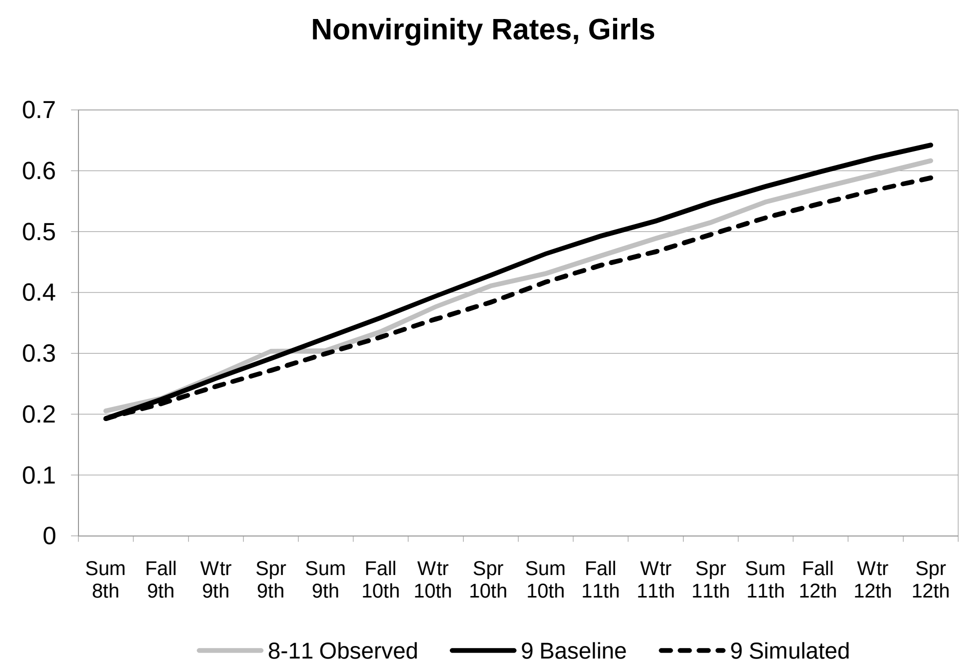


Figure 4b.



NOTES: "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.

Figure 5: Remove Ninth Grade from High School

Figure 5a.

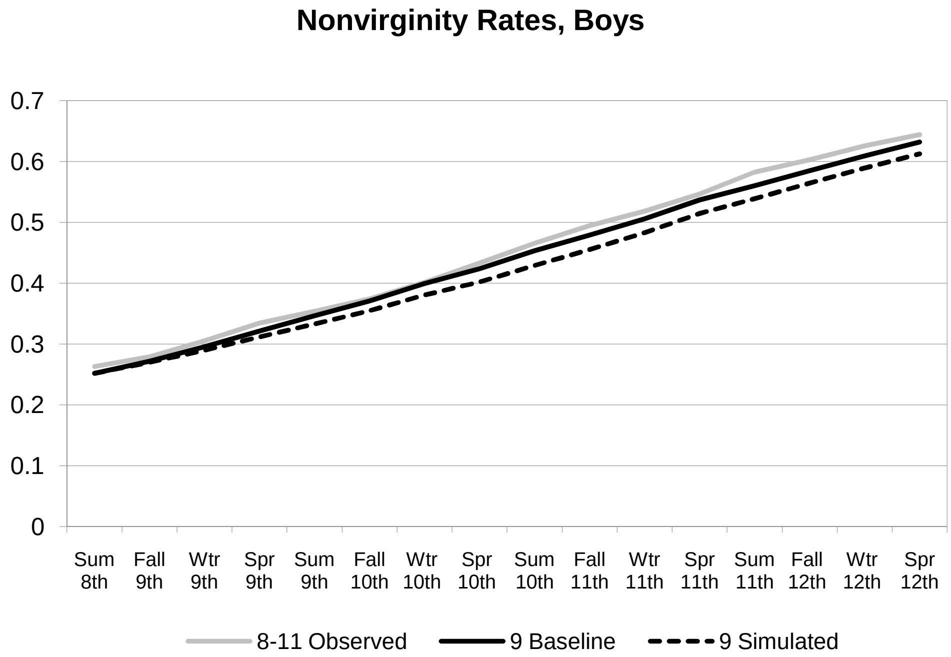
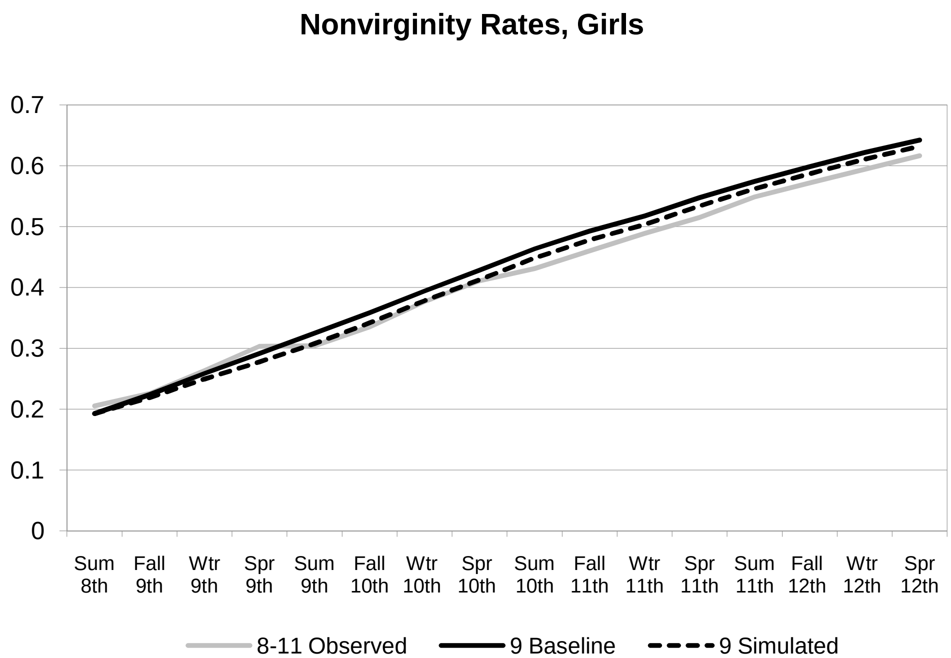


Figure 5b.



NOTES: "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.

Table 1: Descriptive Statistics: Sample Shares with Given Characteristics

Variable	**Unweighted Share**	**Weighted Share**
Individual and family characteristics		
Hispanic	0.179	0.121
Black	0.217	0.163
Younger child	0.500	0.488
Only child	0.190	0.199
Parent with 16+ years of education	0.278	0.270
Family income > $50K	0.239	0.252
Mother married	0.601	0.625
Foreign-born parent	0.156	0.105
Menarche before age 12 (girls)	0.263	0.248
School characteristics		
Urban school district	0.287	0.257
Ninth grade in separate location	0.151	0.065

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	**0.266**	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	**0.154**	0.123	0.144	**0.142**	**0.190**
Girls								
9	**0.212**	**0.207**	**0.208**	0.165	**0.194**	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**

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

Table 3: Fixed Effects 2SLS Estimates of the Composite Effect of Social Interactions

	Boys				Girls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
PANEL A: Reference group is same-gender peers								
Group share nonvirgin	0.168	1.330**	0.744*	0.281	0.656	1.046+	0.956**	0.720*
	(0.377)	(0.401)	(0.292)	(0.471)	(0.505)	(0.543)	(0.371)	(0.320)
Overid. test	6.38	5.43	6.33	2.41	3.78	5.05	8.24	1.93
p-value	0.50	0.61	0.50	0.93	0.88	0.75	0.41	0.98
PANEL B: Reference group is combined peer and supply groups								
Group share nonvirgin	0.962+	2.651**	1.801*	1.965*	-0.404	0.135	1.455	1.431*
	(0.518)	(0.579)	(0.825)	(0.999)	(0.729)	(0.847)	(0.938)	(0.631)
Overid. test	4.07	3.41	11.50	5.24	7.73	7.95	10.55	3.82
p-value	0.85	0.91	0.18	0.73	0.46	0.44	0.23	0.87

Robust standard errors in parentheses. Each column is a separate regression (i.e., estimated separately by grade), thus school-by-grade fixed effects are accomplished with school fixed effects in these regressions.
+ significant at 10%; * significant at 5%; ** significant at 1%

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 (boys / girls)	0.260	0.022
	(0.254)	(0.157)
λ_{13} : above / 2 above (boys / girls)	0.254	0.197
	(0.183)	(0.154)
Chi-square test of (λ_{11},λ_{12},λ_{13},)		
test statistic	6.99	2.32
(p-value)	(0.072)	(0.509)

(continues next page)

Table 4. (continued)

Parameter	Boys	Girls
(continued)		
Type values		
ω^L	-0.270	-0.287
	(0.089)	(0.053)
ω^H	-0.107	-0.089
	(0.066)	(0.056)
Terminal values		
$v(\omega^L)$	-1.608	-0.156
	(0.806)	(0.443)
$v(\omega^H)$	-0.142	0.722
	(0.880)	(0.938)
Type probabilities (π^H)		
Constant term	0.625	-0.369
	(0.857)	(0.531)
Y_0 : 9th grade	0.596	1.765
own gender	(2.438)	(1.635)
Y_0 : 9th grade	-0.291	2.633
opposite gender	(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: Type Distribution

	Boys	Girls
Probability of high type among virgins at the beginning of ninth grade	0.55	0.43
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

Table 6: Probability of Search among Virgins, by Type, and Marginal Effects of Lagged Peer Nonvirginity Rates

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

Notes: Weighted averages combine the search probability for each type weighted by probability that an individual is of each type. Marginal peer effects show the weighted averages of the marginal effects for each type.

Table 7: Average Arrival Rates, and Marginal Effects of Search Behavior among the Opposite Gender

Grade	Boys					Girls			
	Arrival Rate	**Marginal Supply Effects** Same grade	Grade below	Grade above		**Arrival Rate**	**Marginal Supply Effects** Same grade	One grade above	Two grades above
9^{th}	0.103	0.051	NA	0.024		0.108	0.020	0.002	0.019
10^{th}	0.101	0.050	0.024	0.023		0.106	0.020	0.002	0.019
11^{th}	0.106	0.052	0.025	0.024		0.094	0.018	0.002	NA
12^{th}	0.099	0.050	0.023	NA		0.097	0.018	NA	NA

Table 8: Approximation of Equilibrium Beliefs (nonvirginity rates)

	Boys			**Girls**	
	Observed	Simulated		Observed	Simulated
Grade intercepts:					
9th grade	0.014	0.019		0.022	0.024
	(0.003)	(0.003)		(0.004)	(0.004)
10th grade	0.013	0.022		0.023	0.022
	(0.004)	(0.004)		(0.005)	(0.004)
11th grade	0.014	0.028		0.018	0.017
	(0.005)	(0.005)		(0.005)	(0.005)
12th grade	0.015	0.023		0.024	0.017
	(0.005)	(0.005)		(0.004)	(0.005)
Peer group nonvirginity rate:					
Linear term	1.048	1.019		1.047	1.050
	(0.014)	(0.015)		(0.015)	(0.016)
Squared term	-0.089	-0.050		-0.088	-0.086
	(0.014)	(0.015)		(0.016)	(0.015)
Supply groups nonvirginity rates					
Group 1	0.020	0.008		0.012	0.014
(same grade)	(0.006)	(0.006)		(0.007)	(0.007)
Group 2†	0.008	0.007		0.007	0.006
	(0.006)	(0.006)		(0.005)	(0.005)
Group 3†	0.001	-0.001		0.001	-0.002
	(0.004)	(0.004)		(0.006)	(0.005)
R-sq.*	0.965	- - -		0.960	- - -
N	2079	2079		2084	2084
SUR test statistic (obs. vs. sim.)			13.69		
P-value (χ^2, 18 d.f.)			0.75		

Standard errors in parentheses.

* R-squared calculated from separate regressions by gender, with constant term and no 9th grade dummy.

† Group 2 is grade below for boys and grade above for girls. Group 3 is grade above for boys and two above for girls.

Table A-1: 2SLS Estimates for Being Sexually Experienced by Grade, Boys

Reference group:	Same-Gender Peer Group				Combined Peer and Supply Groups			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group								
Peer group	0.168	1.330**	0.744*	0.281	0.962+	2.651**	1.801*	1.965*
	(0.377)	(0.401)	(0.292)	(0.471)	(0.518)	(0.579)	(0.825)	(0.999)
Individual characteristics								
Hispanic	0.171**	0.200**	0.143**	0.074	0.176**	0.197**	0.138**	0.088+
	(0.048)	(0.048)	(0.044)	(0.046)	(0.049)	(0.045)	(0.044)	(0.047)
Black	0.252**	0.213**	0.188**	0.127**	0.260**	0.203**	0.184**	0.128**
	(0.043)	(0.043)	(0.040)	(0.040)	(0.043)	(0.041)	(0.039)	(0.040)
Younger child	0.062*	0.047	0.026	-0.003	0.064*	0.047+	0.032	-0.000
	(0.027)	(0.029)	(0.029)	(0.029)	(0.027)	(0.028)	(0.028)	(0.029)
Only child	0.148**	0.087*	0.064+	0.090**	0.153**	0.085*	0.070*	0.087*
	(0.035)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)	(0.036)
Mother married	-0.120**	-0.092**	-0.074*	0.022	-0.120**	-0.075*	-0.075**	0.025
	(0.030)	(0.031)	(0.029)	(0.028)	(0.030)	(0.030)	(0.029)	(0.028)
Foreign-born parent	-0.072	-0.109*	-0.040	-0.049	-0.067	-0.107**	-0.056	-0.043
	(0.045)	(0.043)	(0.041)	(0.048)	(0.045)	(0.040)	(0.043)	(0.048)
Upper income	-0.016	-0.035	-0.005	0.000	-0.020	-0.044	-0.000	-0.003
	(0.031)	(0.034)	(0.033)	(0.033)	(0.031)	(0.033)	(0.033)	(0.034)
Parental education (4 years college)	-0.106**	-0.105**	-0.081**	-0.080*	-0.103**	-0.106**	-0.084**	-0.081*
	(0.027)	(0.030)	(0.031)	(0.031)	(0.027)	(0.029)	(0.031)	(0.032)
Overidentification test	6.38	5.43	6.33	2.41	4.07	3.41	11.50	5.24
p-value	0.50	0.61	0.50	0.93	0.85	0.91	0.18	0.73
F-stat., first-stage instr.	14.19	19.18	41.68	16.49	40.06	91.51	68.08	18.55
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2241	2713	2873	2908	2241	2713	2873	2908

Robust standard errors in parentheses. Each column is a separate regression (i.e., estimated separately by grade), thus school-by-grade fixed effects are accomplished with school fixed effects in these regressions.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-2: 2SLS Estimates for Being Sexually Experienced by Grade, Girls

Reference group:	Same-Gender Peer Group				Combined Peer and Supply Groups			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group								
Peer group	0.656	1.046+	0.956**	0.720*	-0.404	0.135	1.455	1.431*
	(0.505)	(0.543)	(0.371)	(0.320)	(0.729)	(0.847)	(0.938)	(0.631)
Individual characteristics								
Early menarche	0.065*	0.085**	0.078**	0.094**	0.067*	0.078**	0.084**	0.094**
	(0.027)	(0.030)	(0.029)	(0.026)	(0.026)	(0.028)	(0.028)	(0.026)
Hispanic	0.097*	0.049	-0.003	-0.034	0.095*	0.046	-0.010	-0.035
	(0.046)	(0.047)	(0.045)	(0.041)	(0.045)	(0.044)	(0.043)	(0.042)
Black	0.082+	0.050	0.094*	0.064+	0.076+	0.054	0.096*	0.064+
	(0.045)	(0.040)	(0.039)	(0.037)	(0.044)	(0.038)	(0.038)	(0.038)
Younger child	-0.023	-0.012	-0.001	0.037	-0.019	-0.002	-0.000	0.038
	(0.027)	(0.030)	(0.030)	(0.027)	(0.027)	(0.027)	(0.029)	(0.027)
Only child	0.052	0.073+	0.081*	0.099**	0.050	0.082*	0.085*	0.101**
	(0.036)	(0.038)	(0.037)	(0.031)	(0.035)	(0.035)	(0.036)	(0.031)
Mother married	-0.052+	-0.043	-0.052+	-0.060*	-0.049+	-0.043	-0.050+	-0.060*
	(0.029)	(0.030)	(0.029)	(0.026)	(0.029)	(0.029)	(0.028)	(0.026)
Foreign-born parent	-0.070	-0.029	-0.082+	-0.085*	-0.067	-0.019	-0.081	-0.092*
	(0.045)	(0.051)	(0.049)	(0.041)	(0.042)	(0.047)	(0.049)	(0.042)
Upper income	-0.009	0.008	-0.016	0.010	-0.008	0.005	-0.021	0.006
	(0.030)	(0.033)	(0.034)	(0.032)	(0.029)	(0.031)	(0.033)	(0.032)
Parental education (4 years college)	-0.097**	-0.100**	-0.124**	-0.122**	-0.098**	-0.092**	-0.118**	-0.122**
	(0.026)	(0.031)	(0.033)	(0.031)	(0.026)	(0.030)	(0.033)	(0.032)
Overidentification test	3.78	5.05	8.24	1.93	7.73	7.95	10.55	3.82
p-value	0.88	0.75	0.41	0.98	0.46	0.44	0.23	0.87
F-stat., first-stage instr.	5.34	14.10	20.17	24.07	35.76	26.49	14.04	18.27
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2381	2849	2963	3141	2381	2849	2963	3141

Robust standard errors in parentheses. Each column is a separate regression (i.e., estimated separately by grade), thus school-by-grade fixed effects are accomplished with school fixed effects in these regressions.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-3: Parameter Estimates with Alternative Matching Function

Parameter		**Boys**	**Girls**
Age			
	α	0.098	0.173
		(0.082)	(0.081)
Peer preference interaction			
	γ	0.224	0.213
		(0.080)	(0.082)
Arrival rate			
λ_0 :	9th grade	-2.247	-2.179
		(0.201)	(0.177)
λ_0 :	10th grade	-2.270	-2.144
		(0.150)	(0.149)
λ_0 :	11th grade	-2.211	-2.247
		(0.145)	(0.129)
λ_0 :	12th grade	-2.208	-2.297
		(0.149)	(0.131)
λ_{11} :	same grade	-0.007	0.061
		(0.075)	(0.082)
λ_{12} :	below / above	-0.003	-0.080
	(boys / girls)	(0.023)	(0.085)
λ_{13} :	above / 2 above	0.049	0.069
	(boys / girls)	(0.073)	(0.088)
Chi-square test of (λ_{11},λ_{12},λ_{13},)			
	test statistic	0.47	1.94
	(p-value)	(0.925)	(0.584)

(continues next page)

Table A-3. (continued)

Parameter	Boys	Girls
(continued)		
Type values		
ω^L	-0.301	-0.318
	(0.118)	(0.090)
ω^H	-0.115	-0.037
	(0.074)	(0.089)
Terminal values		
$v(\omega^L)$	-1.556	-0.473
	(0.931)	(0.832)
$v(\omega^H)$	0.021	1.383
	(1.118)	(1.628)
Type probabilities (π^H)		
Constant term	0.154	-0.754
	(0.893)	(0.617)
Y_0 : 9th grade	1.632	1.478
own gender	(2.307)	(1.582)
Y_0 : 9th grade	1.194	3.860
opposite gender	(1.939)	(1.689)
Black	2.774	0.073
	(2.078)	(0.409)
Younger child	0.612	-0.055
	(0.533)	(0.324)
Only child	1.661	2.258
	(1.210)	(1.004)
Parent educ.	-1.937	-1.750
	(1.227)	(0.833)