Spring 5-17-2010

Peer Effects in Sexual Initiation: Separating Social Norms and Partner Supply

Seth O. Richards
University of Pennsylvania, serichar@econ.upenn.edu

Peer Effects in Sexual Initiation: Separating Social Norms and Partner Supply

Abstract
Sexual activity among adolescents is associated with risks such as pregnancy and STDs, and there is substantial policy interest in how peers influence the timing of sexual initiation. This paper measures separate effects for two social mechanisms--peer-group norms and partner availability--using a national sample of high school students. I develop and estimate an equilibrium search and matching model for first sexual encounters that specifies distinct roles for the two mechanisms. Norms are defined based on the share of nonvirgins among same-gender peers, which influences whether an individual searches for a sexual partner. Supply is modeled with an arrival rate for partners, which depends on the search behavior among the opposite gender. The model produces a discrete-time duration to first sex which I estimate with quarterly data on individual virginity status constructed from the Add Health study. The endogeneity of peer behavior with respect to individual behavior is addressed with a combination of strategies. First, I use standard instrumental variables methods to estimate linear regressions for virginity status at the end of each grade. Instruments for group nonvirginity rates are person-specific characteristics such as sibling structure and age of menarche, and the regressions include school-by-grade fixed effects. This analysis demonstrates that school-based social interactions have a large effect on sexual initiation. Second, I estimate the search and matching model via simulated maximum likelihood, in order to decompose this composite effect into separate effects of peer norms and partner availability. Here I control for the endogeneity of peer behavior by (a) defining the norm effect as a function of lagged peer outcomes, (b) including a random effect that is correlated within schools, and (c) using exogenous peer characteristics as supply shifters. I find that peer-group norms have a large effect on the timing of sexual initiation: removing the peer influence on search decisions, 42% fewer boys and 22% fewer girls become sexually active in ninth or tenth grade. Changes in the availability of partners at school (i.e., changes in opposite-gender search behavior) also have a large impact on initiation rates for boys, although not for girls.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Economics

First Advisor
Kenneth I. Wolpin

Second Advisor
David M. Cutler

Third Advisor
Elena Krasnokutskaya & Aureo de Paula

Keywords
peer effects, social networks, adolescents, sexual behavior, search and matching

Subject Categories
Labor Economics

PEER EFFECTS IN SEXUAL INITIATION: SEPARATING SOCIAL NORMS AND PARTNER SUPPLY

Seth O. Richards

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2010

Supervisor of Dissertation

____

Kenneth I. Wolpin

Walter H. and Leonore C. Annenberg Professor in the Social Sciences

Graduate Group Chairperson

____

Dirk Krueger, Professor of Economics

Dissertation Committee

David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University

Elena Krasnokutskaya, Assistant Professor of Economics, University of Pennsylvania

Áureo de Paula, Assistant Professor of Economics, University of Pennsylvania

Peer Effects in Sexual Initiation:

Separating Social Norms and Partner Supply

ACKNOWLEDGMENTS

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 Guy David, Sara Markowitz, Ellen R. Meara, and Nirav Mehta for helpful comments, and to the National Bureau of Economic Research for data access and computer facilities.

Claire, my wife, has sustained me throughout this process. Her intellectual and emotional support kept me moving forward over the many months, as she generously accommodated a partner full of stresses and anxieties. I am grateful for the morning coffee and all the dinners, the listening and critiquing, and especially her sacrifice of time and energy that made this work possible.

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.

ABSTRACT

PEER EFFECTS IN SEXUAL INITIATION: SEPARATING SOCIAL NORMS AND PARTNER SUPPLY

Seth O. Richards

Kenneth I. Wolpin

Sexual activity among adolescents is associated with risks such as pregnancy and STDs, and there is substantial policy interest in how peers influence the timing of sexual initiation. This paper measures separate effects for two social mechanisms—peer-group norms and partner availability—using a national sample of high school students. I develop and estimate an equilibrium search and matching model for first sexual encounters that specifies distinct roles for the two mechanisms. Norms are defined based on the share of nonvirgins among same-gender peers, which influences whether an individual searches for a sexual partner. Supply is modeled with an arrival rate for partners, which depends on the search behavior among the opposite gender. The model produces a discrete-time duration to first sex which I estimate with quarterly data on individual virginity status constructed from the Add Health study. The endogeneity of peer behavior with respect to individual behavior is addressed with a combination of strategies. First, I use standard instrumental variables methods to estimate linear regressions for virginity status at the end of each grade. Instruments for group nonvirginity rates are person-specific characteristics such as sibling structure and age of menarche, and the regressions include school-by-grade fixed effects. This analysis demonstrates that school-based social interactions have a large effect on sexual initiation. Second, I estimate the search and matching model via simulated maximum likelihood, in order to decompose this composite effect into separate effects of peer norms and partner availability. Here I control for the endogeneity of peer behavior by (a) defining the norm effect as a function of lagged peer outcomes, (b) including a random effect that is correlated within schools, and (c) using exogenous peer characteristics as supply shifters. I find that peer-group norms have a large effect on the timing of sexual initiation: removing the peer influence on search decisions, 42% fewer boys and 22% fewer girls become sexually active in ninth or tenth grade. Changes in the availability of partners at school (i.e., changes in opposite-gender search behavior) also have a large impact on initiation rates for boys, although not for girls.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

About one-half of the students in grades nine through twelve in the United States are sexually experienced.[1] Sexual activity begins at young ages, with 14.6 percent of boys and 13.0 percent of girls initiating before age 15, and even higher rates in certain subgroups.[2] The prevalence of sexual activity among adolescents raises substantial concerns, largely because of associated risks such as unplanned pregnancy and sexually transmitted disease.

The initiation of sex is a significant event in itself due to the persistence of sexual activity once started (Arcidiacono, Khwaja, and Ouyang, 2009). A variety of interventions that attempt to delay first sex have been proposed and implemented. Drawing on evidence in psychology and sociology that adolescents are strongly influenced by peer norms, many interventions include an educational program against group norms that promote sex.[3] Another policy targeting a social mechanism is to restrict the supply of partners, who are often met at school. Single-sex schools rep-

[1]Sex is defined as sexual intercourse. This figure comes from the the Youth Risk Behavior Survey.

[2]For example, among non-Hispanic blacks 29% of boys and 23% of girls report sexual intercourse before age 15 (published data from the 2002 National Survey of Family Growth).

[3]Manlove, Romano-Papillo, and Ikramullah (2004) describe several interventions that explicitly target social norms or so-called "peer pressure."

resent one way to do this, but a less drastic option is to isolate ninth grade from the older grades in high school.[4]

Existing econometric work has found large peer effects in adolescent risk 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). These results suggest that some kind of intervention targeting social interactions would be effective. However the methods used in these studies cannot distinguish among the different mechanisms that are relevant to interventions such as the two above, because what they measure is a *composite* effect of social interactions. This effect is defined 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. Such an effect can result from many underlying mechanisms, whereas the effectiveness of an intervention depends on the particular mechanism it targets. This is especially important in the case of sexual initiation because there are (at least) two plausible social mechanisms: social norms among peers and the supply of partners at school. The assessment of these mechanisms is complicated because the level of sexual activity among peers could have both a direct effect on an individual's desire to have sex and an indirect effect via the availability of partners in equilibrium.

In this paper, I apply a model for sexual initiation that provides distinct roles for peer norms and partner availability, in order to measure separate effects for these two mechanisms. This responds to suggestions by Manski (2000), writing on social interactions, and Gruber (2001), writing on youth risk behaviors, to integrate formal economic models with empirical work in order to clarify mechanisms. The starting point for my model is to consider a market for sexual partners defined within the

[4] In fact, some school districts keep the ninth grade in middle school.

student body at a 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. I use a search and matching framework, so this "demand" is modeled as the decision to search for a sexual partner. The probability of finding a partner then depends on the search decisions of others. Thus, the effect of partner availability appears through the changes in the probability of finding a partner due to changes in the search behavior among the opposite gender. Peer norms, on the other hand, affect preferences—i.e., the expected utility of sex that determines the search decision—where "peers" are defined as persons of the same gender, and "norms" is a reasonable interpretation for the influence of peer-group nonvirginity rates on individual tastes.

The empirical analysis has two complementary components. First I estimate the composite effect of social interactions at school using an instrumental variables (IV) method that is standard in this literature. I am able to improve on the existing work on sexual initiation because the data allow me to estimate an effect on virginity status at each grade, and to include school-by-grade fixed effects. Then, having established the presence of a composite effect in my data using standard methods, I structurally estimate the search and matching model to measure separate effects for peer norms and partner availability. This estimation uses simulated maximum likelihood.

The data 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 allows the search and matching model to be estimated as a dynamic process, and it makes possible an innovative econometric strategy to identify the endogenous effects of social interactions.

The results 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 ninth or tenth grade falls by 0.08 (42% of the total) for boys and 0.06 (22%) 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, a one-standard-deviation increase in the share of girls searching for a sexual partner raises the probability of finding a partner each period by 18% for boys in the tenth grade. Overall, these results are consistent with the IV estimation, which finds large composite effects of social interactions.

Policy simulations show that an educational intervention reducing the influence of peer norms would have a larger impact than isolating the ninth grade from older grades to restrict the supply of partners. A program cutting the measured effects of peer norms in half would decrease sexual initiation during the ninth and tenth grades by 27 percent for boys and by 12 percent for girls. On the other hand, removing the ninth grade from high school would decrease initiation in that year by about 13 percent for both boys and girls, but the effect dissipates over time.

In addition to measuring separate effects for peer norms and partner availability in adolescent sexual initiation, this work relates to the broader empirical literature on social interactions where the endogeneity of peer behavior is a central problem for estimation. In an overview of this literature, Moffit (2001) describes three conceptually distinct sources of endogeneity bias: the simultaneity of observed actions, which Manski (1993) calls the "reflection" problem; the correlation of omitted variables among peers; and selection into peer groups.

The most common strategy to address these problems is to use instrumental variables (IV) to provide exogenous variation in peer behavior.[5] Typically, instruments are person-specific characteristics hypothesized not to have a direct effect on the outcomes of others. However, the exclusion restrictions assumed for identification are violated if the distribution of these characteristics among peers has a direct effect on the individual, sometimes called a "contextual" effect (Manski, 1993; Brock and Durlauf, 2001). This can occur if, for example, peers are defined spatially, and mean peer characteristics (especially socioeconomic attributes) relate to local factors that are unobserved. The use of school-by-grade fixed effects in my IV estimation is thus important to remove any such contextual effects that are time-invariant. This strategy, which identifies an effect of social interactions using differences between cohorts that pass through the same grade, is relatively novel in economic work on adolescent risk behaviors.

Apart from IV methods, several recent papers use longitudinal data to address the multiple sources of endogeneity.[6] De Paula (2009) presents a nonparametric test for social interactions based on the simultaneous occurrence of outcomes among peers. Alternatively, if social interactions take time to have an effect, the use of lagged peer outcomes can remove the simultaneity problem (Manski, 1993). However unlike standard models, serial correlation in the individual errors would lead to biased estimates in such models. This is because lagged peer outcomes are themselves affected by the individual's behavior in an earlier period.[7] To account for this and the other sources of endogeneity, some recent papers take advantage of data where individuals appear

[5]For example, among the econometric papers on adolescent risk behaviors referenced earlier, all use IV except for Clark and Lohéac (2007).

[6]The literature includes other methods that do not rely on longitudinal data. Evans, Oates, and Schwab (1992) and Krauth (2006) assume a multivariate normal distribution for the errors and use this to correct for selection or to allow for correlated unobservables (with additional assumptions). Glaeser, Sacerdote, and Scheinkman (1996) assume a specific network structure for individual interactions, which allows them to recover a peer effect from excess variance in aggregate outcomes. Sacerdote (2001) and Katz, Kling, and Liebman (2001) use random assignment to peer groups.

[7]Clark and Lohéac (2007) argue that serial correlation is not important in their setting.

in multiple peer groups by including individual fixed effects (Hanushek et al., 2003; Mas and Moretti, 2009; Arcidiacono, Foster, and Kinsler, 2009; Jackson and Bruegmann, 2009). This strategy removes any permanent component of the unobservable, which addresses endogeneity bias due to both serial correlation in the individual errors and any common omitted variables among peers—assuming these factors are time-invariant.

In my structural estimation, I use similar strategies based on longitudinal data. First, I specify the peer influence on preferences to be a function of lagged peer outcomes. This removes the problem of simultaneity, and it is consistent with the development of social norms over time rather than jointly occurring opportunities to have sex. Then to account for serial correlation in the individual errors, the model includes a random effect as a permanent component of preferences. The distribution of this random effect depends on individual characteristics, and on the shares of non-virgins in each grade at the time when the individual enters high school (to capture initial conditions). This produces a correlation in the unobservable among peers, which addresses correlated omitted variables that are time-invariant. The combination of these two strategies—the use of lagged peer outcomes and a random effect that is correlated among peers—is novel in the literature on social interactions, and it enables me to exploit time-series variation in data where individuals do not appear in multiple peer groups.

There are two further strategies that do not rely on longitudinal data. Person-specific characteristics among the opposite gender play a role similar to IV. They affect preferences, so they function as exogenous predictors of search behavior (i.e., supply shifters). And finally, I use the grade in school to define peers, rather than an endogenous social group like sport teams or nominated friends, in order to avoid the problem of peer group selection.[8]

Apart from the methods used to address endogeneity bias, another innovation in this work relates to how I solve for equilibrium behavior. To determine optimal search decisions, agents need beliefs about the shares of nonvirgins by gender and grade in future periods because these shares affect future payoffs and arrival rates. In order to construct equilibrium beliefs, I apply an insight from recent work on the estimation of discrete dynamic games which uses observed outcome probabilities directly for rational beliefs (Bajari, Benkard, and Levin, 2007; Pakes, Ostrovsky, and Berry, 2007).[9] I adapt this technique to be feasible in my context, where the large (but finite) number of agents produces a complicated distribution of outcomes, by approximating the evolution of nonvirginity rates as an autoregressive process. Obtaining beliefs directly from the data in this way greatly simplifies and speeds the estimation procedure because there is no need to solve for a new equilibrium with each set of candidate parameters.

The rest of this work is organized as follows. The next chapter gives further background on teenage sexual activity and summarizes existing evidence on peer and other influences in this behavior. Chapter 3 presents the model. Chapter 4 describes the data, and chapter 5 contains the IV analysis that demonstrates a composite effect of social interactions. Chapter 6 describes the structural estimation procedure and explains how the endogeneity of peer behavior is addressed in that approach. Chapter 7 gives the results from the search and matching model that measures separate effects for peer norms and partner availability, and chapter 8 presents policy simulations and other counterfactual experiments.

[8]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.

[9]These papers build on the method originally developed for individual dynamic models by Hotz and Miller (1993).

Chapter 2

Background on Adolescent Sexual Behavior in the U.S.

The national Youth Risk Behavior Survey (YRBS), a survey of students in grades 9-12 in the United States administered every two years since 1991, shows that the share of high school students who are sexually experienced has ranged from 46 percent to 54 percent over the past two decades. The share decreased until 2001, and since then has had a small but statistically insignificant increase.[1] For girls the decrease occurred mainly in younger grades, with the nonvirginity rate among ninth graders falling from 32 percent in 1993 (the first year reported by grade) to 27 percent in 2007 while staying at 66 percent among twelfth graders. For boys the decline in sexual experience occurred in all grades (CDC, 1995 and 2008a). The National Survey of Family Growth (NSFG) shows simliar patterns by age rather than grade. From 1995 to 2002, nonvirginity rates among (never married) 15 to 17 year-olds decreased from 38 to 30 percent for girls and from 43 to 31 percent for boys. The rate among 18 and

[1] 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.

19 year-olds remained statistically unchanged at 68 percent for girls, while it dropped from 75 to 64 percent for boys (Abma et al., 2004).

Published data from the NSFG also provide information on the frequency of sex and number of partners for teenagers (age 15-19).[2] Sexual activity is highly persistent: 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. This persistence indicates the importance of initiation. Moreover, the decrease in sexual *experience* from 1995 to 2002 almost completely accounts for the decrease in sexual *activity* over that period.[3] The rates of activity and experience fell by the same relative amounts, meaning that the probability of being active conditional on being experienced was unchanged.[4] The modal number of sexual partners per year, among sexually active teenagers, is one: 47 (61) percent of nonvirgin boys (girls) have one partner over a 12-month period, while only 8 percent of either gender have four or more partners. On the other hand, most individuals who become sexually active as teenagers have more than one partner during their teenage years. Among sexually experienced 18 and 19 year-olds, 71 percent of boys and 65 percent of girls report more than one partner in their lifetimes.

Among the risks associated with adolescent sexual behavior, the greatest attention has gone to unplanned pregnancy. In his 1995 State of the Union address, President Bill Clinton labeled the "epidemic" of teenage pregnancy and out-of-wedlock childbearing "our most serious social problem."[5] The rate of childbearing among women

[2]The figures that follow are calculated from data published in Abma et al. (2004).

[3]Sexual activity is typically defined as intercourse within the past three months.

[4]For boys, the rates of sexual experience and sexual activity both fell by 17%; for girls, the rate of sexual experience fell by 8% and the rate of sexual activity fell by 9%.

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

aged 15-19 in the U.S. is substantially higher than in other developed nations, despite declines from a peak in the early 1990s. Increased contraceptive use accounted for most of this decline, but corresponding with the large drop in nonvirginity rates at younger ages, reductions in sexual activity accounted for 23 percent of the decrease in childbearing among girls aged 15 to 17 from 1995 to 2002 (Santelli et al., 2007).

Teenage childbearing is associated with negative outcomes for both the mothers and their children. For example, teenage mothers have lower educational attainment, and their sons are more likely to be incarcerated at some point in their lives (Hoffman, 2006). There are also large public expenditures on the children of teenage mothers, such as an estimated $1.9 billion for medical care and $2.3 billion in foster care per year (Hoffman, 2006).

Another risk associated with sexual activity is sexually transmitted disease (STD). The prevalence of STDs is higher among teenagers and 20-24 year-olds than any other age groups. For the two most common and well-reported STDs, chlamydia and gonorrhea, there were nearly 480,000 cases among teenagers in 2007 (CDC, 2008b). Considering these and six other major STDs, Chesson et al. (2004) estimate that the lifetime medical cost of treating the amount 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 small literature in economics examines adolescent sexual behavior, and most of this work considers individual-based, as opposed to social, factors. In a review, Levine (2001) describes a process with three sequential choices: whether to have sex, whether to use birth control, and then if pregnant, whether to have an abortion. He explains that most of the work in economics at that point focused on fertility and collapsed the three choices into a single, determinstic decision to give birth.[6] Levine (2001) then examines the effect of indirect costs associated with pregnancy and unprotected sex, such as welfare benefits for children and the prevalence of AIDS. He finds that statewide variations in these costs over time have an impact on sexual activity and contraceptive use. Other recent papers have studied the relationship between substance use and sexual behavior (Rees, Argys, and Averett, 2001; Sen, 2002; Markowitz, Kaestner, and Grossman, 2005). The evidence is mixed: for example, Markowitz, Kaestner, and Grossman (2005) find no effect of alcohol or marijuana use on sexual activity using individual fixed effects or IV, but they do find an effect on condom use as well as a reduced form relationship between state alcohol policies and STD rates.

In addition to this work, Oettinger (1999), Walker (2003), and Arcidiacono, Khwaja, and Ouyang (2009) present models of the decisions to have sex and use contraception, which highlight specific aspects of these behaviors. Oettinger (1999) shows how sex education could either increase or decrease sexual initiation by changing the expected utility of sex and unplanned pregnancy, or changing the (real or perceived) risk of pregnancy conditional on sex. He then estimates that sex education increases the hazard rate of initiation for girls, controlling for common unobservables among siblings. Walker (2003) focuses on perceived pregnancy risk, which in his model influences contraceptive effort. He presents survey data on subjective expectations which show that most teenagers have accurate beliefs about their own risk of pregnancy, except for sexually experienced girls in poor families, who underestimate their risk.

[6] Levine (2001) notes two papers that considered the impact of AIDS on condom use, although they did not analyze teenagers.

Arcidiacono et al. (2009) estimate a dynamic model with joint decisions on having sex and using contraception, which includes utility terms to capture habit persistence in sexual activity. They find large fixed and transition costs that lead to persistent behavior.

As for social influences, Fletcher (2007) is the only econometric work estimating an effect of social interactions on sexual activity per se. Case and Katz (1991) and Evans, Oates, and Schwab (1992) estimate social effects on teenage childbearing, with the former using average teen fertility in a neighborhood and the latter using the share of economically disadvantaged students at school as explanatory variables.

Outside economics, large literatures in psychology and sociology analyze the role of peer norms in adolescent behavior, and there is specific evidence on sexual initiation. Kinsman et al. (1998), Santelli et al. (2004), and Sieving et al. (2006) measure peer norms through self-reported individual perceptions about the level of sexual activity among peers, peer attitudes toward sex, and the social gains for becoming sexually active. All three studies find that norms defined this way have a significant association with the probability of initiating sex. Moreover, this work specifically asks for perceptions about how many peers are *already* sexually experienced, which supports the use of lagged peer outcomes in my model. In addition, earlier studies surveyed by the National Research Council (NRC) Panel on Adolescent Pregnancy and Childbearing (1987) indicate that the relevant norms operate within gender.

Related to this area of research, many interventions to delay first sex target peer norms. Two examples are “Safer Choices,” first implemented in 1993 with 2,000 ninth and tenth grade students in ten high schools in California and Texas, and “Draw the Line/Respect the Line,” (DL/RL) first implemented in 1997 with 1,500 middle school students in California (Manlove, Romano-Papillo, and Ikramullah, 2004). These programs consist of about 20 classroom sessions spaced out over two or three school years, which include learning about social norms and practicing communication skills. To give a flavor of exactly how these programs work, Session 2 in Safer Choices is described as

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

Similarly, Session 5 in DL/RL "discusses the role that friends play in respecting the line. Role play scenarios are used to practice showing respect for another person's limits" (p. 21).[7]

In addition to peers defined as friends, researchers have examined the influence of siblings and romantic partners on sexual initiation. Rodgers et al. (1992), Widmer (1997), and Argys, Rees, Averett, and Witoonchart (2006) give evidence that individuals with older siblings tend to initiate sex earlier. Kaestle, Morisky, and Wiley (2002) find that girls with romantic partners who are much older have a higher probability of initiating sex.

Finally, a number of individual and parental characteristics have been shown to predict early sexual initiation. The NRC Panel on Adolescent Pregnancy and Childbearing (1987, chapter 4) provides a useful overview. Black race and low socioeconomic status meaured by parental education or family income are strong predictors, although some recent research indicates the race effect is mostly among 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).

[7]These programs were evaluated by comparing outcomes across schools that were randomly assigned into treatment or control groups. The results found that DL/RL 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).

Chapter 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 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).[1] 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 become sexually active.

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, the match probabilities among individuals 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.

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

3.1 Model Specification

The model applies to individuals, i, located in a local market for partners, m, who have a gender, $\tau \in \{b, g\}$.[2] Virginity status at the start of a period is denoted $y_{i,t-1}$, with 0 meaning virginity. Each period virgins make a search decision $d_{it} \in \{0, 1\}$, 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 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 *share* of searchers among the opposite gender at the school. The arrival rate does not depend on the search behavior of the same gender because there is no a constraint on the number of partners, so there is no competition between same-gender individuals. The share of searchers among the opposite gender at i's school is denoted N_{it}, and this includes the behavior of both virgins and nonvirgins. Because there is also an external market for partners, even with zero searchers at school the arrival rate is positive.

The arrival rate is given by a gender and age specific function, which is specified as a logit:

$$\lambda_{it} = \lambda_{a_{it}}(N_{it}) \equiv \frac{\exp(\lambda_{0a_{it}} + \lambda_1 N_{it})}{1 + \exp(\lambda_{0a_{it}} + \lambda_1 N_{it})} \tag{3.1}$$

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

(with subscripts τ_i suppressed). The main reason to allow the parameters λ_{0a} to vary with age is to capture changes in the amount of contact with the external market as students progress to older grades. The key parameter of interest in (3.1) is λ_1, which gives the effect of partner availability at school.

Individuals derive utility from their virginity status. The per-period payoff for being sexually experienced is a linear combination of age, the share of peers who are already nonvirgins, denoted $Y_{i,t-1}$, a permanent individual component, ω_i, and iid preference shocks, ϵ_{it}. Peers are individuals of the same gender in the same grade as i. The per-period utility for a nonvirigin is then is expressed as

$$u(a_{it}, Y_{i,t-1}, \omega_i, \epsilon_{it}) \equiv \overbrace{\alpha a_{it} + \gamma Y_{i,t-1}}^{u_{it}} + \omega_i + \epsilon_{it}. \tag{3.2}$$

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

The term $\gamma Y_{i,t-1}$ represents the effect of peer norms. Stated more precisely, the model relates lagged peer outcomes to the expected utility of sex. I interpret this as an effect of social norms, based on the evidence from sociological work on sexual initiation. 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) expresses 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 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, nontrivial terminal values are needed. For nonvirgins, I eliminate the peer influence on preferences after high school (there is no further data, anyway) and hold the age and 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.[3]

I express lifetime utility using the Bellman representation, with value functions denoted $V_a(y_{i,t-1}, Y_{t-1}, \omega_i, \epsilon_{it})$. The vector Y_{t-1} (8×1) contains the shares of nonvirgins 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. $V_a(1, Y_{t-1}, \omega_i, \epsilon_{it})$ gives the expected lifetime utility for a nonvirgin, which has an analytical expression:

$$u(a_{it}, Y_{i,t-1}, \omega_i, \epsilon_{it}) + \sum_{s=a_{it}+1}^{A} \beta^{s-a_{it}} \; \mathrm{E}_t u(s, Y_{i,s-1}, \omega_i, \epsilon_{is}) + \beta^{A-a_{it}} \frac{\alpha A + \omega_i}{1-\beta}, \tag{3.3}$$

where E_t denotes the individual's expectation given his or her information set (and recall that $a_{it} = t$ for the reference cohort).

For a virgin, the value function is a more complicated object that incorporates the search decision and the arrival rate. It is expressed as

$$\begin{aligned}
&V_{a_{it}}(0, Y_{t-1}, \omega_i, \epsilon_{it}) = \max_{d_{it}} \\
&d_{it} \; \mathrm{E}_t\Big[\lambda_{it} \cdot \Big(u_{it} + \omega_i + \epsilon_{it} + \beta V_{a_{it}+1}(1, Y_t, \omega_i, \epsilon_{i,t+1})\Big) + (1-\lambda_{it}) \cdot \beta V_{a_{it}+1}(0, Y_t, \omega_i, \epsilon_{i,t+1})\Big] \\
&+ (1-d_{it}) \; \beta \mathrm{E}_t V_{a_{it}+1}(0, Y_t, \omega_i, \epsilon_{i,t+1}),
\end{aligned} \tag{3.4}$$

where u_{it} is defined in (3.2). The second line above expresses that an individual who

[3]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 the expected value of any preference interactions in the future.

searches ($d_{it} = 1$) will become a nonvirgin with probability λ_{it} and will remain a virgin with probability $(1 - \lambda_{it})$. The third line is the value of not searching, in which case the individual advances to the next period still a virgin.

To form the expectations in (3.3) and (3.4), individuals need beliefs over the sequence of shares of nonvirgins among peers $(Y_{it}, Y_{i,t+1}, \dots)$ and the arrival rate (λ_{it}). 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 rules 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 rules over the distributions of ω_j and ϵ_{jt}, and 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 for N_{it}, the share of searchers, which in turn gives the distribution of λ_{it}. How I implement this is explained in section 3.2.

I use an approximation to fully rational beliefs about the evolution of Y_t, 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:

$$\mathrm{E}[Y_{kt}|Y_{k,t-1}] = \psi_{0k} + \psi_1 Y_{k,t-1} + \psi_2 Y_{k,t-1}^2 + \sum_{j \in s(k)} \psi_{3j} Y_{j,t-1}. \tag{3.5}$$

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 vector autoregression that stacks these elements is denoted $\psi(Y_{t-1})$. As in Krusell and Smith (1998) and Lee and Wolpin (2006), this first-order autoregression fits the true evolution of the aggregate state extremely well. There are two details in the implementation of these beliefs. First, because school populations are finite, the approximation incorporates the impact of an individual's choice and outcome on his or her own group's share of nonvirgins.[4] Second, because the aggregate state does not contain information on younger cohorts below high school, the nonvirginity rates for each new cohort of ninth graders are predicted based on the previous cohort.[5]

Finally, the expected costs and benefits of sexual activity embodied in the permanent individual term ω_i may relate to the probability of initiation prior to 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_i 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.[6] There are two reasons to think that the distribution of ω_i 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 share of nonvirgins among peers) indicates a higher ω_i for the individual. Second, if ω is uncorrelated but there are common opportunities to initiate sex prior to ninth grade, the distribution of ω_i among the remaining virgins is affected by selection.

In addition to Y_0, the conditional distribution of ω_i also depends on a vector of exogenous, permanent individual characteristics, x_i. 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, so that $\omega_i \in \{\omega^k\}_{k=1}^{\kappa}$, which assumes

[4]There is a straightforward modification to (3.5) to account for a known value of y_{it} in Y_{it}, with a group of given size n_i. The approximation ignores any impact on other groups.

[5]I use the nonvirginity rate of one cohort in the summer after ninth grade ($t = 4$) to predict the rate for the new cohort in the same period. I do this by inverting the following regression for the annual growth of nonvirginity rates during ninth grade: $\mathrm{E}Y_{k4} = \Pi_0 + \Pi_1 Y_{k0}$. The formula for the prediction is then $Y_{k'4} = Y_{k4}/\Pi_1 - \Pi_0/\Pi_1$, where subscripts k' and k denote the new and old cohorts.

[6]The vector Y_0 contains all grades in the summer before the reference cohort enters ninth grade, because these nonviriginity rates are needed for the beliefs about $\{Y_t\}_{t=1}^{T}$.

there are κ "types" when it comes to sexual initiation. The conditional distribution of ω_i, which is for virgins at the beginning of ninth grade, is specified as a multinomial logit:

$$\Pr(\omega_i = \omega^k \mid Y_0, x_i) = \pi_{k|Y_0,x_i} \equiv \frac{\exp(\pi_0^k + Y_0'\pi_1^k + x_i'\pi_2^k)}{1 + \sum_{l=2}^{\kappa} \exp(\pi_0^l + Y_0'\pi_1^l + x_i'\pi_2^l)}, \tag{3.6}$$

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

3.2 Individual Behavior and Equilibrium

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 (3.4) to

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

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

$$d_{it} = 1 \text{ iff } u_{it} + \omega_i + \epsilon_{it} + \beta \mathrm{E}_t V_{t+1}(1, Y_t, \omega_i, \epsilon_{i,t+1}) > \beta \mathrm{E}_t V_{t+1}(0, Y_t, \omega_i, \epsilon_{i,t+1}). \tag{3.7}$$

Thus a virgin will search if and only if the value of becoming sexually active exceeds the value of remaining a virgin.[7]

The age-specific value functions for virgins do not have an analytical expression, but they can be numerically constructed by backward recursion starting from the

[7] 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.

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_i from the set of values $\{\omega^k\}$, and sample x_i and the membership of the peer and supply groups from their joint empirical distribution.

To evaluate the value functions in (3.4) 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 (3.7), and integrate over the unobserved values of ω_j and ϵ_{jt} among those virgins. However, the random values of Y_{t-1} in the solution points do not correspond to the observed 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 use 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. To construct the regression coefficients, I start with an initial value which allows 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 (3.4) with this approach, I first 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.

To solve for the equilibrium, I need equilibrium beliefs about the evolution of Y_t. To construct these beliefs I estimate the autoregression ψ in (3.5) directly from the observed aggregate data, in a preliminary stage before the estimation of the structural parameters. As in Bajari, Benkard, and Levin (2007) and Pakes, Ostrovsky, and Berry (2007), this method assumes only one equilibrium is observed, and it assumes a steady state from one cohort to the next. 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 post-estimation by re-estimating ψ on data simulated from the model and comparing the two sets of estimates. Also, because the autoregression fits the observed evolution of Y_t extremely well, with $R^2 > 0.99$, I use a degenerate distribution at the expected values for the beliefs in the approximation. This avoids the need to integrate over a distribution in each future period when solving for individual behavior.

An alternative to the two-stage 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.

Chapter 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 self-directed portion of the survey on a laptop computer at respondents' homes. Included in these questions, respondents are asked if they have ever had sexual intercourse which is defined explicitly.[1] Those who say yes are then asked to report the month of first sex. Both rounds of interviews ask these questions of all respondents,

[1]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.)

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. 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 extimation 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"). Each cohort, which is observed for one or two years, is shown as a separate line positioned over the appropriate ages. The black line then averages among all individuals observed at a given age, to produce a complete path through high school for a synthetic cohort. 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. These graphs show that a large portion of individuals initiate sex during high school. The share of nonvirgins among boys increases from just over 26 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 individual or family characteristics come from Wave I. In the structural estimation of the search and matching model, I use indicators for black race, parental education, and sibling status as the characteristics (x_i) that affect the distribution of the individual preference term (ω_i). The education variable indicates whether one parent has 16 years of education (i.e., is a four-year college graduate). The sibling variables are two dummies that indicate whether the individual is a younger sibling or is an only child. I use these variables because they are predetermined characteristics that have been shown to predict early sexual initiation in other work, and they have clear relationships with the expected costs or benefits of sexual activity. For example, race and parental education relate to expected labor market outcomes, which affect the relative cost of a pregnancy. Younger siblings may learn from older siblings about the benefits of sex or how to reduce the costs (e.g., through birth control). In the IV estimation of the composite effect of social interactions, I use additional exogenous characteristics to improve the power of the instruments, and because the computational burden of additional coefficients is negligible. These variables are indicators for: Hispanic ethnicity; mother currently married; a foreign-born parent; relatively high household income, defined as above $50,000; and early menarche for girls, defined as before the median age of 12.

The shares of individuals with each of these characteristics are shown in table 1. The table presents both the unweighted and weighted shares, and these figures are similar except for black race, Hispanic ethnicity, and a foreign-born parent, which reflect oversamples in the sample design. Table 1 also shows the shares of individuals in urban school districts and in districts where the ninth grade is in a separate location from the rest of the high school. The weighted and unweighted values for the latter are quite different because a large number of individuals are sampled in one high school that only has grades 10 through 12.[2] Either way, however, the vast majority of ninth graders go to school with older students.

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 the second year). Under random sampling, these correlations are equivalent to correlations in virginity status between two individuals, one from each specified group.[3] 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 by grade. Overall, table 2 shows there are large correlations in virginity status within schools, about 0.2 in magnitude. This would produce substantial variation in nonvirginity rates across schools, which is a notable consequence of social interactions, although it can also result from unobserved factors.

For the endogenous supply of partners, only certain grades of the opposite gender are used in the empirical implementation. These grades are shown in bold in table 2: for boys, they 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.[4] 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

[2]In-home interviews, which are the source of my data, were conducted with nearly all students at 16 of the high schools in its sample. One of these is a large, urban high school with only grades 10-12, which is the cause of the discrepancy between weighted and unweighted figures.

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

[4]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.

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. Thus in the estimation, the grade-specific constant term in the arrival function (λ_{0a}) captures the probability of finding a partner either from outside school or in one of these other grades. 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 (table 2).

Chapter 5

Instrumental Variables Analysis

For the first empirical analysis, I apply an IV method that is standard in the literature on peer effects. There are two purposes for this exercise. First, I am able to improve on estimates from previous work due to the richness of my data. I can estimate an effect of social interactions on virginity status at each grade, because virginity is assessed over time and because the clustered sample provides a sufficient number of individuals in each grade per school. In addition, because the data contain multiple cohorts that pass through each grade, I am able to include school-by-grade fixed effects. This controls for any correlated unobservables or contextual effects that are invariant over the observation period. Other recent work on school-based social interactions in adolescent risk behaviors also uses school fixed effects for this purpose (Lundborg, 2006; Clark and Lohéac, 2007; Trogdon, Nonnemaker, and Pais 2008; Fletcher, 2009). However, except for Lundborg (2006), these papers use the same fixed effect for all grades.[1] This removes a common effect on levels but does not address differences across schools in how risky behaviors increase with age. Yet because sexual initiation is a one-time transition, it is natural to think of a duration process and focus on

[1]Lundborg (2006) defines peer groups at the classroom level and uses variations across classrooms within grades, at a single point in time.

how nonvirginity rates rise with age. If unobserved school-wide factors affect the hazard rate, there will larger differences across schools at younger grades and smaller differences at older grades. These are not captured with a single school fixed effect.

The second purpose for the IV analysis is to demonstrate the presence of social interactions in my data using a standard method that has different identifying assumptions than the structural estimation procedure. As in Fletcher (2007), which uses a different dataset but from the same time period, the IV estimation here finds a large composite effect of social interactions on sexual initiation.

5.1 Regression Model

The empirical models in this analysis are linear regressions for virginity status at the spring quarter of each grade in high school. They are based on the following general form:

$$y_{iat} = \pi_{0at} + \pi'_{1a} x_i + \pi_{2a} \bar{y}_{(-i)t} + \pi'_{3a} \bar{x}_{(-i)} + w_{ma} + e_{iat}, \tag{5.1}$$

which is estimated separately by gender and grade (as usual, gender subscripts are suppressed). The term π_{0at} is a time-specific intercept to capture differences in the average outcomes of each cohort. The variables $\bar{y}_{(-i)t}$ and $\bar{x}_{(-i)}$ contain average outcomes and characteristics of the reference group (discussed below). The school-by-grade fixed effect is w_{ma}.

It is important to note that these regressions cannot be interpreted as approximations for the search and matching model. Seen in terms of the model, virginity status at a point in time is the cumulative result of a transition process over many periods. Because the per-period transition probabilities multiply out to produce the probability of being a nonvirgin at a point in time, there is no simple correspondence between the components of the transition process specified in the model and this cumulative outcome.[2] In addition, the search and matching model applies to individuals who are virgins at the beginning of high school, while the IV analysis does not condition on initial status (this is standard). Given these differences, the baseline specification uses *contemporaneous* peer outcomes in the regressions rather than lagged outcomes, because this is the standard formulation in the literature. Regressions using lagged outcomes produce similar results.

Another difference with the model is the use of a single reference group (this also follows the literature); the regressions do not distinguish between a within-gender effect and a cross-gender effect. Having separate effects in a static regression model would be misleading because both coefficients would reflect combinations of the peer influence on demand and the effect of partner availability. To see this, consider an exogenous increase in the share of (same gender) peers who are sexually experienced. This would directly increase an individual's demand for sex. However it would also indirectly increase the supply of partners in equilibrium because the increase in peer nonvirginity would raise the arrival rate experienced by the opposite gender, thereby increasing nonvirginity among the opposite gender which in turn raises the arrival rate for the individual. This indirect effect becomes clear when followed over time: in period 1, the nonvirginity rate among peers exogenously increases; the search behavior of these nonvirgins then raises the arrival rate of partners for the opposite gender, so in period 2 their nonvirginity rate increases as well; hence in period 3 both the individual's search probability (demand) and arrival rate (supply) are higher. In a static framework, these effects cannot be disentangled. This is different than a shock to supply that is common among peers, which can be addressed with proper instruments. The difference is that both the direct effect on demand and the indirect

[2] To see this, suppose that the per-period transition probability for individual i is P_{it}. Then the probability that i is nonvirgin at time t, conditional on being a virgin at time 0, is $1-\prod_{s=1}^{t}(1-P_{is})$.

effect on supply are consequences of the exogenous increase in peer nonvirginity, due to the equilibrium nature of the model.

To explore different sources of variation in the data, I estimate regressions with two alternative constructions of the reference group: 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. The regression coefficient for either construction combines the demand and supply effects contained in the model, but the exact combination would be different for the two constructions.

To identify (5.1), I must assume that $\pi_{3a}^{k} = 0$ for some elements of $\bar{x}_{(-i)}$; in other words, there are valid exclusion restrictions that allow certain individual characteristics to be used as instruments. I first estimate regressions that assume all of $\pi_{3a} = 0$, so that the means of all characteristics among the reference group serve as instruments. I then estimate regressions where π_{3a}^{k} is nonzero for socioeconomic characteristics which are likely to have contextual effects. Here only the sibling indicators and the early menarche variable are excluded. Finally, I estimate regressions that include the school-by-grade fixed effects (w_{ma}). For these I use all the potential instruments ($\pi_{3a} = 0$), in order to have as much "within" variation as possible for the first stage. Here the effect of social interactions is identified from differences between the mean characteristics of the two cohorts that are observed at each grade level, which predict different group nonvirginity rates. The validity of the instruments in this case rests on the assumption that any contextual effects they may have do not vary over a two-year period.

I test the exclusion restrictions in these various specifications with an overidentification test. This requires that at least one element of $\bar{x}_{(-i)}$ can be excluded from (5.1), which is itself an untestable assumption.[3] In the fixed-effects IV regressions,

[3]This test also assumes that any potential violations of the specification do not cancel each other

this assumption pertains to the difference in mean characteristics from one year to the next. All of the characteristics are predetermined, so it seems reasonable that this should hold for at least one of them.

5.2 Estimates of the Composite Effect of Social Interactions

Tables 3 and 4 present estimates of the composite effect of social interactions (also referred to as the "social effect") from several different specifications, using the alternative exclusion restrictions and reference groups described above. Full results for each of these specifications are contained in appendix tables A1-A4. The coefficients reported in tables 3 and 4 give the effect of the share of nonvirgins in the reference group on the probability that an individual is sexually experienced by the spring quarter of the indicated grades. For example, in panel A, column 11, of tables 3 and 4, the point estimates imply that a 10 percentage-point increase in the share of nonvirgins among same-gender peers increases the probability of being sexually experienced by 7.4 percentage points for boys and 9.6 points for girls in the eleventh grade.

These results provide strong evidence of a school-based social effect on sexual initiation for both boys and girls. Although overidentification tests indicate the exclusion restrictions may be violated without fixed effects, the fixed effects estimates also find large effects of social interactions for both genders in most grades. Also it is interesting to note that differences in the estimates across specifications using the same reference group—i.e., when contextual effects or school-by-grade fixed effects are included—typically are not statistically significant.

out in the test statistic, which is possible due to the loss of degrees of freedom that results from estimating the coefficients. See Hayashi (2000, p. 218).

The use of school-by-grade fixed effects appears to be a successful strategy for removing any contextual effects that would otherwise invalidate the instruments. The overidentification tests in columns 9-12 of tables 3 and 4 have high p-values, which indicates the instruments are properly excluded from the second stage. As expected, the standard errors are much larger when the fixed effects are included, but there is still enough variation in the data to yield statistically signficant results in most grades when both constructions of the reference group are considered.[4] Even with these fixed effects the instruments retain their predictive power, as indiciated by the first-stage F-statistics (on the instruments) which are shown in tables A1-A4. On the other hand, adding contextual effects for mean socioeconomic characteristics among the reference group appears to help with instrument validity for girls in some cases (table 4, column 7 vs. column 3) but not for boys (table 3, columns 5-7).

Although the fixed effects estimates are noisy, there are suggestive differences across grades and between boys and girls. For boys, the effect of nonvirginity rates among peers (i.e., same gender, same grade) appears strongest in grades 10 and 11. For girls there is no such difference across grades. When the reference group includes the opposite gender as well as the peer group, the fixed effects estimates indicate a larger social effect for boys than for girls, especially in grades 9 and 10. This is interesting in light of the structural estimates, which find an effect of opposite gender search behavior on the arrival rate for boys but not for girls.

Tables 5 and 6 show estimates of the social effect from fixed effects regressions that only contain the smaller set of variables (race, parental education, and sibling indicators) used in the structural estimation. Although there is no direct link between the coefficients in these regressions and the parameters of the model, these estimates

[4]The estimates are particularly noisy using the combined peer and supply group, which makes sense because the supply groups overlap for the two cohorts observed in each grade, so there is less "within" variation in the outcomes and characteristics of the reference groups.

demonstrate the robustness of the IV results using only the variation and controls provided by these limited characteristics. For comparison, columns 1-4 of these tables repeat the estimates from the full specifications (i.e., columns 9-12 of tables 3 and 4). Columns 5-8 show results with the smaller set of variables using contemporaneous group outcomes, and columns 9-12 show results using lagged outcomes. Both sets of estimates are noisier than the richer specification, but still the qualitative patterns are similar and so are the magnitudes in many cases.

Finally, because individual characteristics have different associations with sexual initiation depending on gender, one might think of using separate variables for the mean characteristics of boys and girls as instruments. However, the mean characteristics of boys and girls at a school are highly correlated, which can be a problem for estimation. Appendix table A5 shows that the estimated social effects using the combined peer and supply groups instrumented in this way are similar to those instrumented with a single mean for each characteristic (as in panel B of tables 3 and 4). But in results not shown, the gender-specific instruments often have opposite signs and relatively large magnitudes in the first stage, which indicates a multicollinearity problem.

Chapter 6

Structural Estimation

I use simulated maximum likelihood to estimate the search and matching model. The model produces a discrete-time duration to first sex; accordingly, the likelihood contributions have a general form that is common for duration analysis. The search decisions of virgins are unobserved, so the per-period transition probability is given by the product of the probabilities of searching and finding a partner. The likelihood function also contains observations on nonvirgins, which takes advantage of data on the arrival of subsequent partners in order to improve the estimation of the arrival rate.

6.1 Likelihood Function

The likelihood contributions for virgins express the probability for the observed durations to first sex. They 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 (3.7) is satisfied.[1] With ϵ distributed as a standard normal, its CDF denoted Φ, this product is

$$L_{it}(\omega) \equiv \Phi\left[u_{it} + \omega + \beta \mathrm{E}_t V_{a_{it}+1}(1, Y_t, \omega, \epsilon_{i,t+1}) - \beta \mathrm{E}_t V_{a_{it}+1}(0, Y_t, \omega, \epsilon_{i,t+1})\right] \cdot \int \lambda_{a_{it}}(N_{it})\ f(N_{it})\ \mathrm{d}N_{it}. \tag{6.1}$$

The solution to the model for a particular set of parameters provides the expected future values of virginity and nonvirginity inside Φ, so the value of this probability can be calculated. Simulation is needed for the arrival rate, in order to integrate over the unobserved search decisions of opposite-gender virgins which generate N_{it}.

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

$$L_i(\omega) \equiv L_{it_i^*}(\omega) \cdot \prod_{t=1}^{t_i^*-1} [1 - L_{it}(\omega)]. \tag{6.2}$$

Finally, the likelihood contribution for the duration to first sex is the combination of these type-specific probabilities, weighted by the probability of each type given initial conditions:

$$\mathcal{L}_i \equiv \sum_{k=1}^{\kappa} \pi_{k|Y_{m0},x_i} L_i(\omega^k). \tag{6.3}$$

For individuals who are not observed to initiate sex, there is a standard modification to (6.2).

I use simulation to construct the expected arrival rate in (6.1), because the search decisions of virgins in the supply groups are unobserved and depend on individual shocks. For each simulation round, $r \in 1 \ldots R$, I simulate a search decision, d_{jt}^r, for each virgin, j, in the three supply groups. This proceeds by drawing ω_j^r from the

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

appropriate distribution and then comparing the type-specific search probability for individual j, given by $\Phi[\dots]$ in (6.1), 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_{it}^r. Then averaging $\lambda_{a_{it}}(N_{it}^r)$ across simulation rounds produces an approximation for the expected arrival rate.

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 rate function also appears in the likelihood contributions for the durations to first sex, which use the arrival of the first sexual partner.

To construct a likelihood contribution for the arrival of second partners, I define z_{it} to indicate that sex with a second partner has been reported in the current or some previous period. Then the contribution can be expressed as

$$\mathcal{A}_i \equiv \prod_{t=1}^{T} \left(\left\{ (1 - \mathrm{E}[\lambda_{a_{it}}(N_{it})])^{1-z_{it}} \, \mathrm{E}[\lambda_{a_{it}}(N_{it})]^{z_{it}} \right\}^{1-z_{i,t-1}} \right)^{y_{it}} . \tag{6.4}$$

The terms inside the curly brackets give the probability for a discrete-time duration, and they only affect the value of the expression from the first period when $y_{it} = 1$ (first partner) through the first period when $z_{it} = 1$ (second partner).[2] 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 an independent event and is independent of individual characteristics, the likelihood contributions in (6.4) simply multiply with the likelihood contributions in (6.3). 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 $t = 0$.

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 approximation is calculated via numerical differentiation of the individual likelihood contributions.

6.2 The Endogeneity of Peer Behavior

The structural estimation strategy addresses the multiple potential sources of endogeneity of peer behavior as follows. First, the use of lagged peer outcomes in the search decision removes the simultaneity problem. This addresses simultaneity both among same-gender peers and between genders; for the latter, it is because the arrival rate is determined by opposite-gender search behavior, which is also a function of lagged outcomes. Second, the presence of a permanent unobservable (ω) addresses the problem with serial correlation in the unobservable that arises in a dynamic model of social interactions. Third, ω is correlated within schools, which controls for corre-

[2] 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.

lated omitted variables that are time invariant. This combination of lagged outcomes and a permanent unobservable that is correlated among peers, while natural for a dynamic model, appears to be novel in the literature on peer effects. Finally, the definition of peer groups is intended to avoid selection bias. Peer groups are defined by grade, and it seems unlikely that individuals would systematically skip or repeat grades in order to affect their chances of sexual initiation.

Under this strategy, the variation used to identify the peer influence on search decisions comes from differences in Y_{it} across peer groups conditional on Y_{m0}. Two groups with the same Y_0 would have the same distribution of ω for their members (controlling for individual characteristics), but then different values of Y_{it} in later periods would produce different search probabilities. Thus the strategy directly uses the kind of variation that motivates interest in social interactions—the magnification of small differences given similar initial conditions. This variation also identifies the effect of partner availability, because the arrival rate depends largely on the shares of nonvirgins among the opposite gender (i.e., the opposite gender elements of the vector Y_{mt}).[3]

6.3 Left-Censored Observations

Although the start of the duration period is known (i.e., birth), the presence of unobserved heterogeneity presents a problem when individuals are first observed after ninth grade. The problem is that individuals who are still virgins in later grades are more likely to have low values of ω compared with virgins at the beginning of high

[3]The current implementation uses only three elements of Y_{m0} to condition the distribution of ω: the ninth, tenth, and eleventh grade shares for the individual's own gender. This restriction is useful to reduce the number of parameters and to avoid using too many highly correlated variables. But given the argument above, a revised implementation will add the ninth grade share for the opposite gender.

school. The estimation procedure needs to account for this, or else the duration dependence captured by the age parameter α would have a negative bias. Moreover, because the hazard rate is a function of time-varying arguments, there is not a simple way to integrate over the unobserved periods.

To update the distribution of ω for a virgin who is first observed after ninth grade, I create approximate, type-specific hazard rates and use these to calculate the probability, for each type, of being a virgin when the individual is first observed. These rates are approximated with data from the younger cohorts at the individual's school, which relies on a steady state from one cohort to the next.[4] To make the approximation, I regress the type-specific transition probabilties from the likelihood function ($L_{it}(\omega)$) on relevant state variables which include the lagged shares of nonvirgins by gender and grade ($Y_{m,t-1}$). I then use these regressions to predict type-specific hazard rates for the individual before the observation period, when he or she was in a younger grade in high school. For 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 I can then use Bayes Rule to update the initial distribution of ω for the individual.

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).

[4]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.

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

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

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

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).

Chapter 7

Results from the Search and Matching Model

The estimated model fits the observed patterns in sexual intiation, 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 to the end of senior year. 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 virgins in 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 forward projection fits the observed nonvirginity rates of older cohorts in upper grades, which indicates the successful use of multiple, overlapping cohorts to estimate a complete path through high school.

The structural parameters are shown in table 7. The effect of lagged peer nonvirginity rates on expected utility, γ, is large and strongly significant for boys. For girls this parameter is about two-thirds as large, and is marginally significant with a z-statistic of 1.57 (p-value = 0.12). The age parameter, α, has the opposite pattern for boys and girls. This suggests that girls are more influenced by individual development than by peer interactions, relative to boys.

The effect of opposite-gender search behavior on the arrival rate is shown 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 chapter 4. These three parameters are jointly significant for boys (chi-square, 3 d.f. = 10.4, p-value = 0.02), but not for girls. For boys the estimated effect of opposite-gender search behavior is largest from girls in the same grade, while for girls the effect is twice as large from boys who are two years older compared with boys in the same grade, although these differences are not well measured.

Finally, two permanent types are sufficient to fit the observed growth of nonvirginity rates during high school, as well as differences in the age trends along various observable characteristics (not shown).[1] I refer to these as "low" and "high" types. Both types have negative values for ω, although the difference from zero is not statistically significant for the high type. These parameters indicate similar expected costs and benefits of sex for boys and girls: the chi-squared test statistics (1 d.f.) for the gender differences in ω^L and ω^H are 0.04 and 0.15 respectively, with p-values above 50%. The type-specific terminal values for virgins, $\nu(\omega)$, are poorly measured, which is not surprising given the simple linear growth in nonvirginity rates.

To provide better interpretation for the estimated model, tables 8 and 9 present average search decisions and arrival rates by gender and grade, as well as the marginal effects of lagged peer nonvirginity rates on search probabilities (in table 8) and of

[1] Arcidiacono et al. (2009) also find that two types are sufficient in their work on adolescent sexual behavior.

current supply group search behavior on arrival rates (in table 9). Table 8 pertains to the search decisions by virgins. For both boys and girls, the average probability of search for low types is fairly small in ninth grade (about 0.15) but increases throughout high school to about 0.5 in twelfth grade. High types start in ninth grade with about a 0.5 and 0.7 probability of search, respectively, for boys and girls, and this increases more for girls than boys as individuals age. High-type girls who are still virgins in the twelfth grade are very likely to be searching, according to the model. Averaging between the two types, weighted by the probability of being each type, the probability of search among virgins is between 6 and 9 percentage points higher for girls than for boys, depending on the grade. This fits with the faster growth in nonvirginity rates among girls during high school.

At each grade, the marginal effect of lagged nonvirginity rates among peers is about double for boys what it is for girls. These marginal effects are substantial in relation to the average search probabilities, especially in younger grades. For example, they imply that an increase of one standard deviation (about 0.17) in the nonvirginity rate among peers would raise the probability of search by 0.07 (0.04) for boys (girls) in the ninth grade, which is 21% (10%) of the average search probability in that grade.

Table 9 shows that the arrival rate of partners is similar for boys and girls and is fairly constant across grades. Because the constant term λ_0 varies by grade, the arrival rate does not automatically increase over time even though nonvirginity rates do. Also, not all grades have three supply groups available at school; for example, there are no older groups for twelfth graders. 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 of search behavior among the opposite gender are listed for each of the supply groups present. Compared with the marginal effect of peer outcomes on search decisions, the raw marginal effects here are an order of magnitude smaller for boys. Still, the combined effect of search behavior among the supply groups can be large. 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 statistically significant so the marginal effects may only reflect sampling noise.

Table 10a shows the probability of being high type among individuals who are virgins at the beginning of each grade. The share of high types among the remaining virgins decreases with grade because high types are more likely to initiate sex, leaving more low types among the uninitiated. Table 10b gives the average change in the probability of being high type attributable to each of the four permanent characteristics used in the estimation. This is calculated among virgins at the start of ninth grade, which is the population for which the conditional distribution of ω is specified. Each row shows the difference in the high-type probability ($\pi_{H|Y_{m0},x_i}$) produced by changing the designated indicator variable from 0 to 1, with all the other variables kept as they are observed, and averaged over this sample. These "effects" are qualitiatively similar to the coefficients on these variables estimated in the IV analysis (shown in appendix tables A1-A4).

Chapter 8

Policy Simulations and Other Counterfactuals

Figures 3 through 8 present the results of policy simulations and other counterfactuals that demonstrate features of the model. 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 predicted"); 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.

For 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 with the new beliefs. I iterate until the parameters in the beliefs approximation converge, which usually occurs in fewer than 10 iterations.

The counterfactuals in figures 3 and 4 function as decompositions, to demonstrate the overall impact of peer-group norms and partner availability on sexual initiation during high school. Figure 3 eliminates the effect of peer norms by setting $\gamma = 0$. The results indicate that the effect of peer norms on the timing of sexual initiation is about twice as large for boys as for girls, throughout high school. Also, norms have a larger relative impact in younger grades: in the simulation, the share of individuals who initiate sex in ninth or tenth grade falls by 0.08 (42% of the total) for boys and 0.06 (22%) for girls, while the share who initiate at any point during high school falls by 0.09 (24%) for boys and 0.06 (13%) for girls.

Figure 4 eliminates the effect of opposite-gender search behavior on the arrival rate, by setting $\lambda_1 = 0$. This is purely a decomposition, because individuals would likely compensate for the absence of available partners at school by increasing their search efforts elsewhere. The simulation shows that changes in the search behavior of boys at school have very little effect on the initiation rate for girls, while the availability of girls at school does impact boys. 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 is reduced by one third.

Figure 5 simulates an educational program that cuts the effect of peer norms in half, by replacing the estimated γ with $\gamma/2$. This is intended to predict the impact of interventions against peer pressure, such as the two described in chapter 2, on a national population. The simulated program affects both boys and girls. Hence the reductions in sexual initiation incorporate both the decrease in demand for sexual partners that is directely due to the diminished peer influence, and the decrease in supply that is an equilibrium effect of the change in the strength of peer norms among the opposite gender. The simulated program has a larger relative impact in younger years, decreasing sexual initiation during the ninth and tenth grades by 27% for boys and 12% for girls.

Figures 6 simulates isolating the ninth grade from the rest of high school. In the model, this is accomplished by setting the parameters for older supply groups in the arrival rate to zero for ninth graders (i.e., $\lambda_{g12}, \lambda_{g13}$, and λ_{b13}) and setting the parameter for the younger supply group to zero for tenth grade boys (λ_{b12}). This intervention decreases initiation in the ninth grade by about 13% 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 effect on tenth graders and the stronger peer influence among boys that multiplies the initial policy impact.

Figure 7 presents a simulation that removes all the nonvirgins from high school (in all grades) when the ninth grade cohort enters. This is intended to show the importance of initial conditions, both within a peer group and from the supply groups. The excluded nonvirgins are still counted toward the simulated nonvirginity rates presented in the graphs, so the basis for comparison is the same. The results looks similar to the simulation with the peer influence cut in half, in figure 5, and in fact the effect of removing nonvirgins works mostly within peer groups. Simulations that remove only the nonvirgins of one gender (not shown) indicate little effect on the opposite gender. This is probably due to the fact that the arrival rate is a function of the *share* of searchers among the opposite gender, and the share of virgins who search at the beginning of high school is reasonably large (see table 8).

Figure 8 simulates the effect of "virginity pledges" on non-pledgers.[1] This works by setting the value of ω^L (for low types) to a sufficiently negative value so that low-type individuals have a negligible probability of search.[2] These individuals represent the

[1]See Bearman and Bruckner (2001) or Rosenbaum (2009) for evidence on the outcomes of pledgers themselves.

[2]A value of -1 is sufficient to keep the search probability below 0.01 through the end of high school.

pledgers. The graphs show the nonvirginity rates among high-type individuals, who represent non-pledgers. The population in these graphs is restricted to individuals who are virgins at the start of high school, because total nonvirginity rates are not a good basis for comparison given that low-type individuals do not initiate sex. There is little effect on high-type girls, because they have such a high probability of search in the estimated model (see table 8), and the supply of boys has little effect on their initiation rates (see figure 4). For high-type boys on the other hand, the amount of initiation during high school is reduced by 10%. This is an effect of both smaller nonvirginity rates among peers and a smaller share of searchers among girls at school. Of course, to the extent that pledgers and non-pledgers are not peers or potential partners for each other, this overstates the spillover effects of virginity pledges.

Chapter 9

Conclusion

In two separate analyses with different identifying assumptions, I find evidence that social interactions in high school have large effects on sexual initiation. This is in line with results from other work on peer effects in adolescent risk behaviors, including earlier evidence on sexual initiation. The magnitude of the composite effect of school-based social interactions is large: at some grade levels in high school, the estimated effect size is close to 1 or even above it. This means that an increase in the share of nonvirgins among an individual's reference group raises his or her probability of being nonvirgin by an equivalent amount, which represents a large multiplier individual behavior. Although this composite effect is estimated with linear regressions, this suggests the possibility of a nonlinear "tipping point," in which small changes in initial conditions or small shocks can lead to large differences across schools.

By applying an economic model of the underlying process, I am able to decompose the composite effect and provide an assessment of two distinct social mechanisms. This improves our understanding of adolescent sexual behavior, and it is useful for policy because the mechanisms—social norms among peers and the supply of partners at school—are susceptible to different interventions. The results indicate that programs counteracting peer norms that promote sexual activity have more potential to delay sexual initiation than do policies separating younger grades from older grades. In addition, boys appear to be more sensitive than girls to the social mechanisms studied in this work, which suggests designing interventions that are targeted to boys.

This work could be extended in a number of ways. First, it would be valuable to conduct the two empirical analyses on middle school students. Although the number of individuals who are sexually active at these ages is smaller, there is much policy concern about sexual initiation before the ninth grade. In addition, work in sociology indicates that peer norms have a large effect at these ages. The data in the Add Health study could be sufficient for this extension, although the sample sizes are smaller and the grade structure of middle schools is more variable which complicates the definition of the market for partners.

Second, future work would benefit from data over time on subjective perceptions of peer sexual behavior and reported intentions to initiate sex.[1] The analysis in this paper assumes that individuals have accurate beliefs about the shares of nonvirgins among their peers. However some survey evidence indicates that adolescents overestimate the share of peers who are sexually active (National Campaign to Prevent Teen Pregnancy, 2004), and it would be interesting to see how beliefs relate to the actual behavior of peers. Reported intentions to initiate sex, if reliable, could be used as observed search decisions. This would make it easier to include heterogeneity in the arrival rate, and to allow for possible competition for partners between individuals of the same gender.

[1]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.

Bibliography

[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): 1-48.

[2] Arcidiacono, P., A. Khwaja, and L. Ouyang. 2009. "Habit Persistence and Teen Sex: Could Increased Access to Contraception have Unintended Consequences for Teen Pregnancies?" Working paper.

[3] Arcidiacono, P., G. Foster, and J. Kinsler. 2009. "Estimating Spillovers using Panel Data, with an Application to the Classroom." Working paper.

[4] 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.

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

[6] Bearman, Peter S. and Hannah Bruckner. 2001. "Promising the Future: Virginity Pledges and First Intercourse." *American Journal of Sociology*, 106(4): 859-912.

[7] Brock, William A. and Steven N. Durlauf. 2001. "Interactions-Based Models." In *Handbook of Econometrics*, Volume 5 ed. J. J. Heckman and E. Leamer, 3297-3380.

[8] 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.

[9] Centers for Disease Control and Prevention. 1995. "Youth Risk Behavior Surveillance: United States, 1993." *Morbidity and Mortality Weekly Report*, 44(No. SS-1).

[10] ——. 2008a. "Youth Risk Behavior Surveillance: United States, 2007." Morbidity and Mortality Weekly Report, 57(No. SS-4).

[11] ——. 2008b. *Sexually Transmitted Disease Surveillance, 2007.* Atlanta, GA: U.S. Department of Health and Human Services.

[12] 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.

[13] 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.

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

[15] 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.

[16] Fletcher, Jason M. 2009. "Social Interactions and Smoking: Evidence using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects." *Health Economics*, forthcoming.

[17] ——. 2007. "Social Multipliers in Sexual Initiation Decisions among U.S. High School Students." *Demography*, 44(2): 373-388.

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

[19] Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics*, 111(2): 507-548.

[20] Gruber, Jonathan. 2001. *Risky Behavior among Youths: An Economic Analysis.* Chicago and London: University of Chicago Press.

[21] 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.

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

[23] 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.

[24] 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*, forthcoming.

[25] Kaestle, Christine E., Donald E. Morisky, and Dorothy J. Wiley. 2002. "Sexual Intercourse and the Age Difference between Adolescent Females and their Romantic Partners." *Perspectives on Sexual and Reproductive Health*, 34(6): 304-309.

[26] Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116(2): 607-654.

[27] 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.

[28] 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.

[29] Krauth, Brian V. 2006. "Simulation-Based Estimation of Peer Effects." *Journal of Econometrics*, 133(1): 243-271.

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

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

[32] 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.

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

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

[35] Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives*, 14(3): 115-136.

[36] ——. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531-542.

[37] Marín, Barbara V., Douglas B. Kirby, Esther S. Hudes, Karin K. Coyle, and Cynthia A. Gmez. 2006. "Boyfriends, Girlfriends and Teenagers' Risk of Sexual Involvement." *Perspectives on Sexual and Reproductive Health*, 38(2): 76-83.

[38] 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*, 95(2): 263-266.

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

[40] 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.

[41] 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.

[42] 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.

[43] National Campaign to Prevent Teen Pregnancy, The. 2004. "American Opinion on Teen Pregnancy and Related Issues, 2003." *Science Says* brief, No. 7.

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

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

[46] 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.

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

[48] 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.

[49] Rodgers, Joseph L., David C. Rowe, and David F. Harris. 1992. "Sibling Differences in Adolescent Sexual Behavior: Inferring Process Models from Family Composition Patterns." *Journal of Marriage and Family*, 54(1): 142-152.

[50] Rosenbaum, Janet E. 2009. "Patient Teenagers? A Comparison of the Sexual Behavior of Virginity Pledgers and Matched Nonpledgers." *Pediatrics*, 123(1): e110-120.

[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] Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116(2): 681-704.

[54] 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.

[55] Santelli, John S., Laura D. Lindberg, Lawrence B. Finer, and Susheela Singh. 2007. "Explaining Recent Declines in Adolescent Pregnancy in the United States: The Contribution of Abstinence and Improved Contraceptive use." *American Journal of Public Health*, 97(1): 150-156.

[56] 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.

[57] 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.

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

[59] Walker, James R. 2003. "Pregnancy and Fertility Expectations: Estimates of Bounded Rationality and Unintended Births." Working paper.

[60] 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.

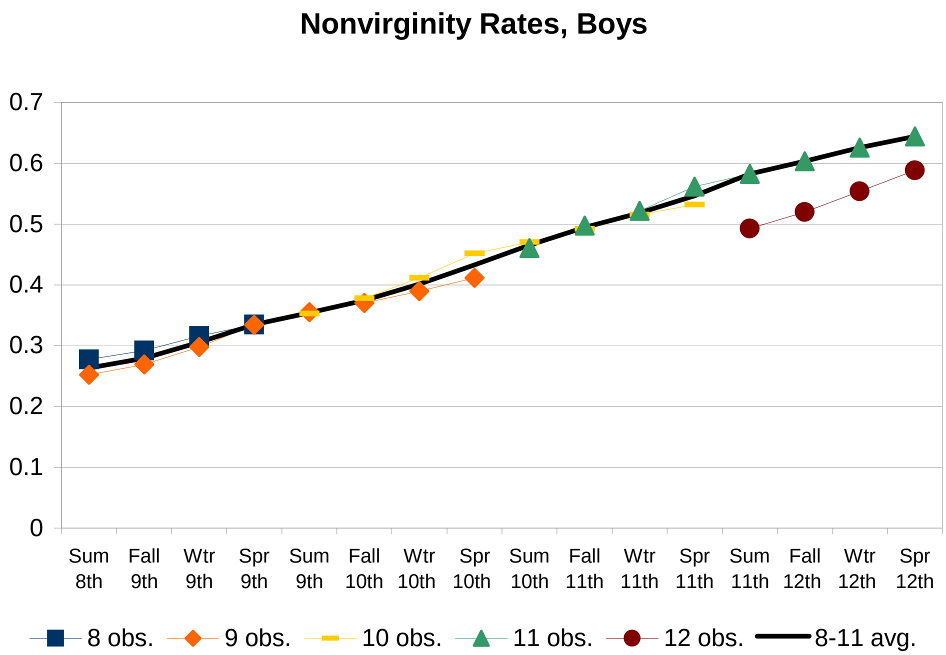
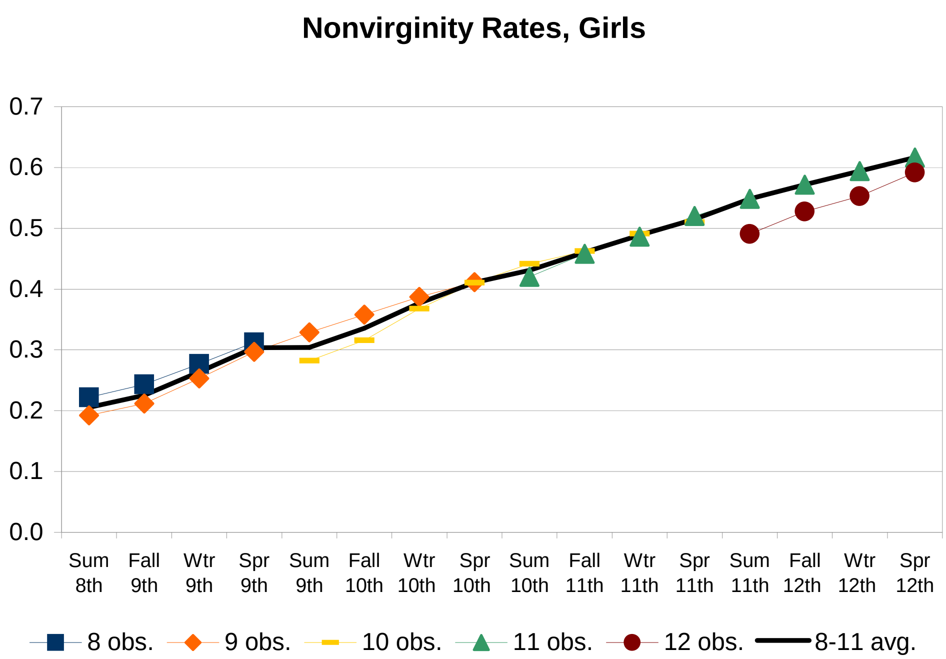


Figure 1b.



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.

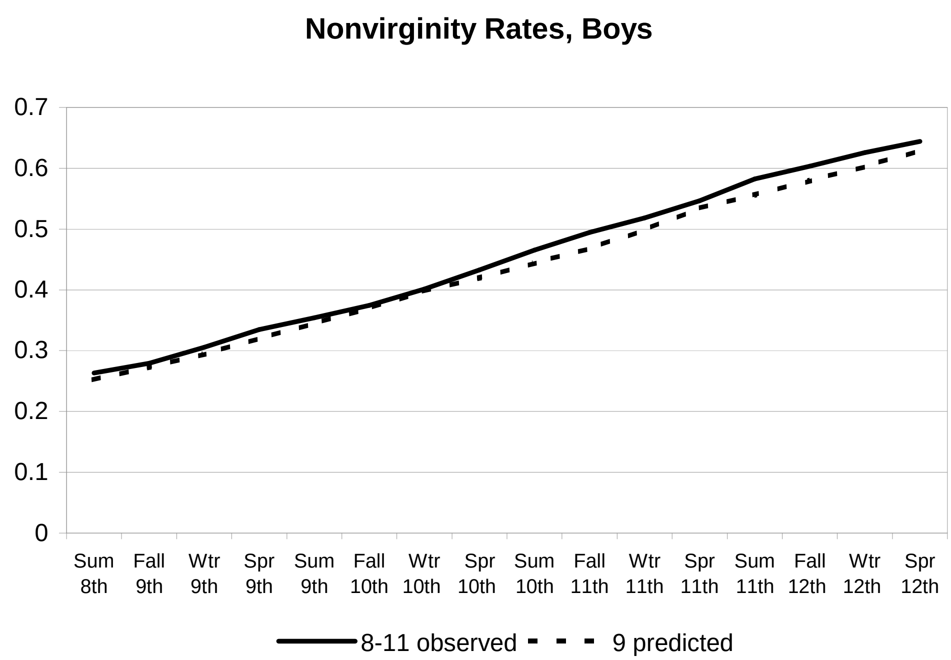
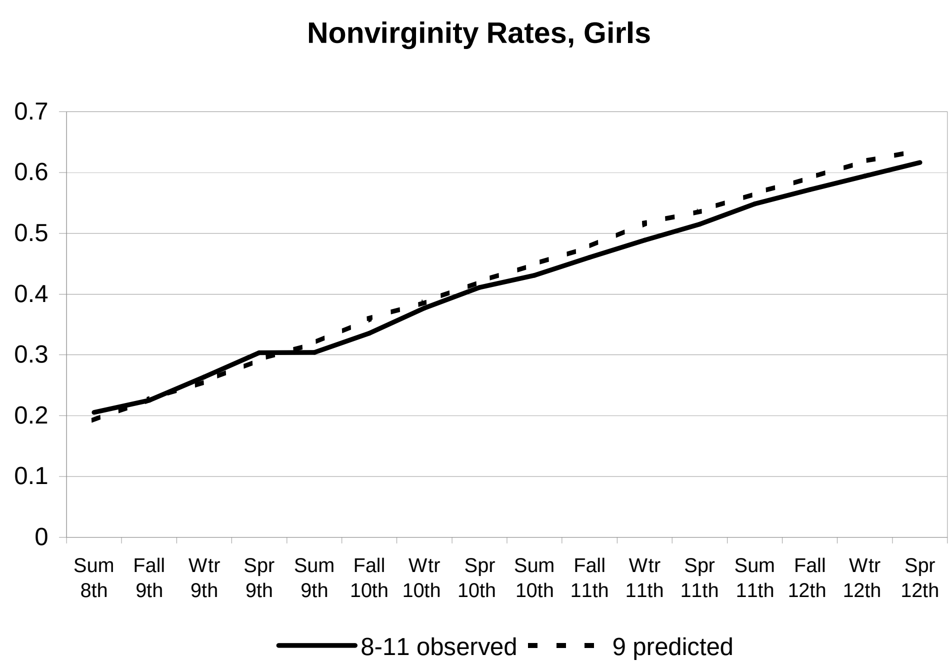


Figure 2b.



NOTES: "8-11 observed" shows the observed rates for the 8th-11th grade cohorts, combined into a synthetic cohort; "9 predicted" shows the prediction for the 9th grade cohort from the estimated model.

Figure 3: Eliminate Effect of Peer Norms on Search Decisions

Figure 3a.

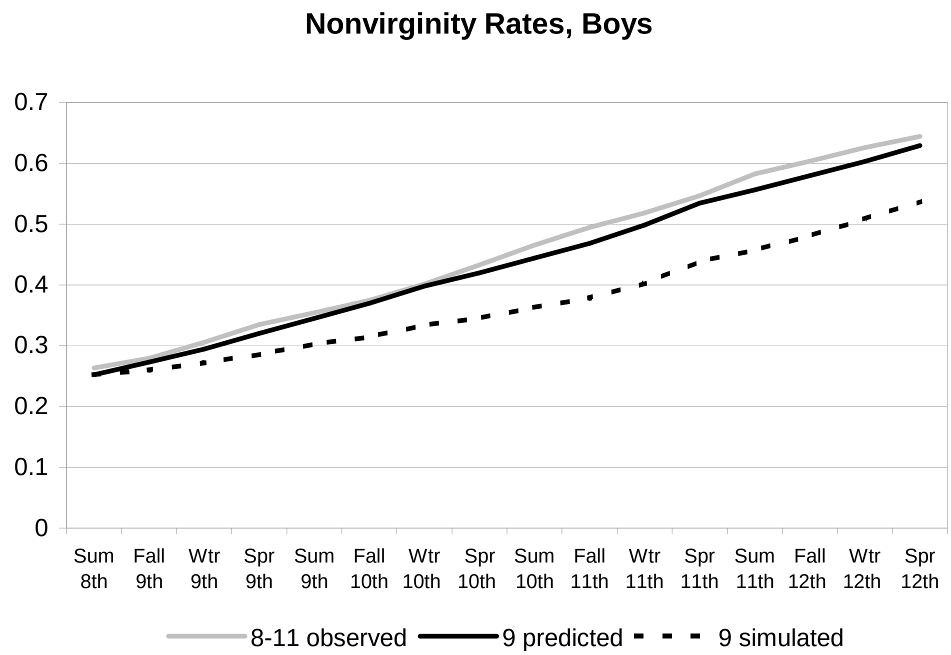
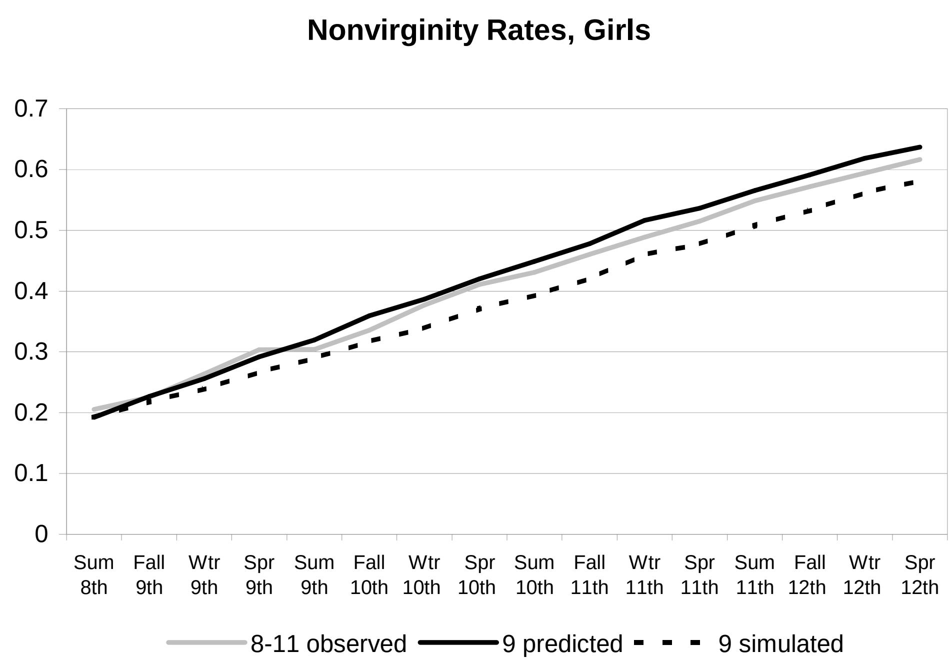


Figure 3b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is prediction from modified model.

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

Figure 4a.

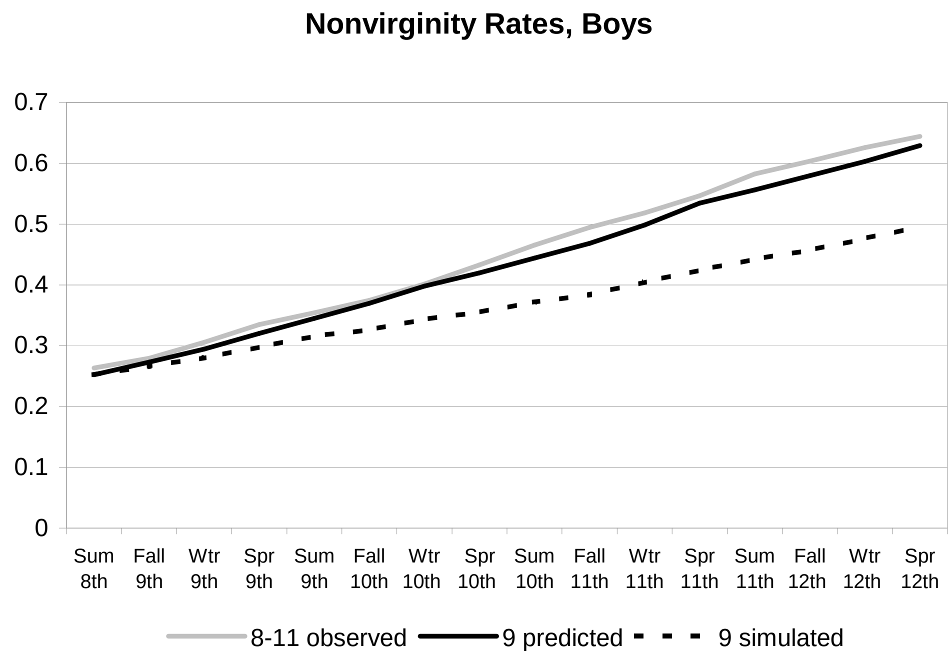
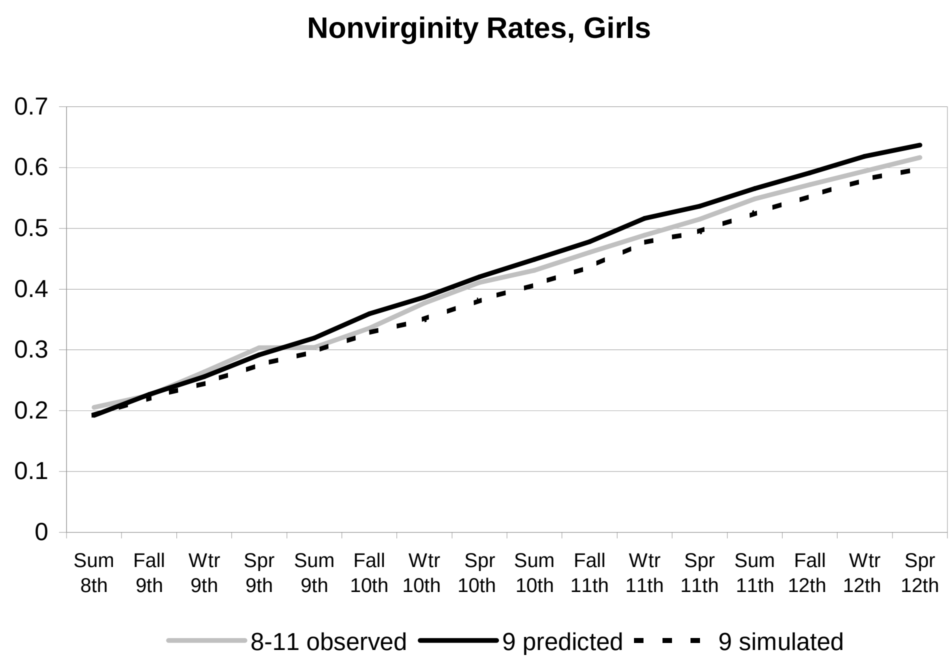


Figure 4b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is prediction from modified model.

Figure 5: Educational Program Cutting Effect of Peer Norms in Half

Figure 5a.

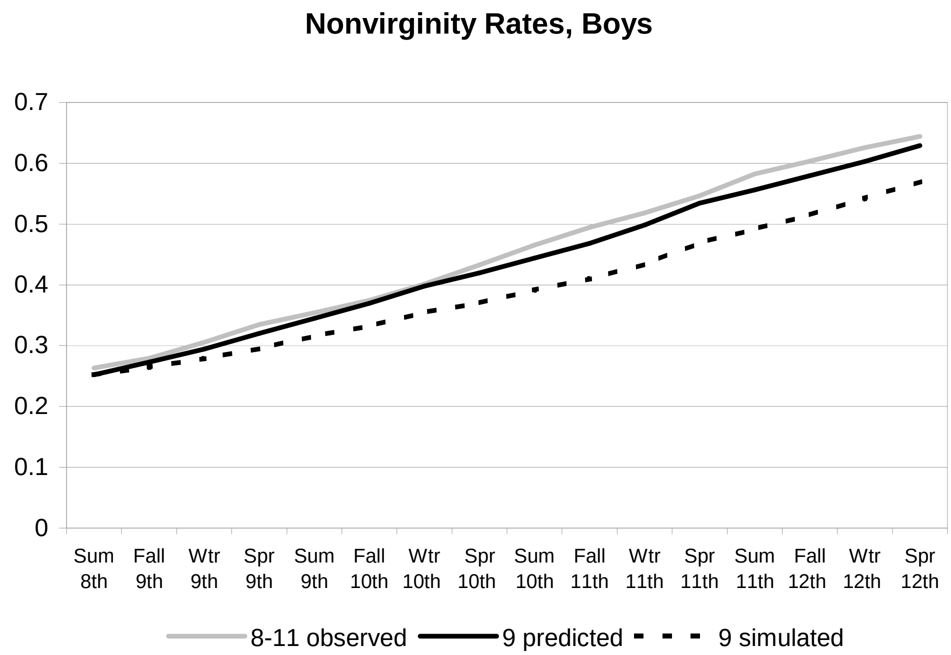
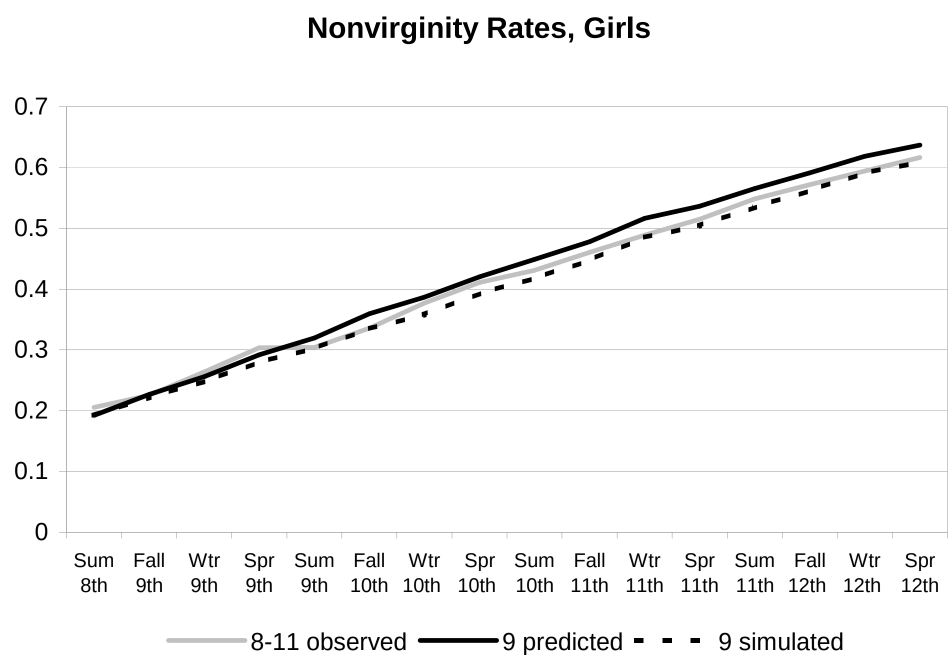


Figure 5b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is prediction from modified model.

Figure 6: Remove Ninth Grade from High School

Figure 6a.

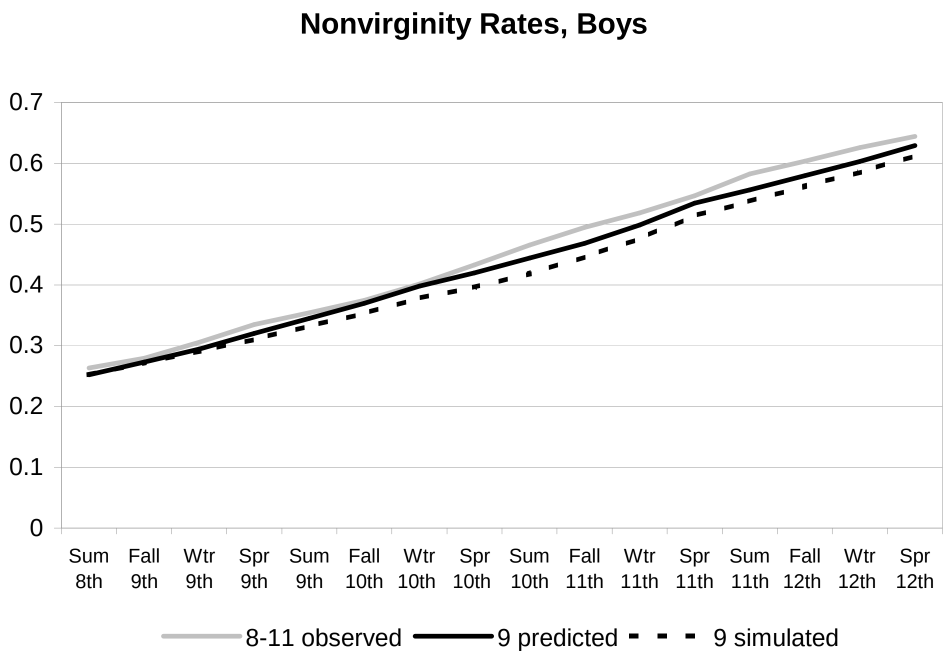
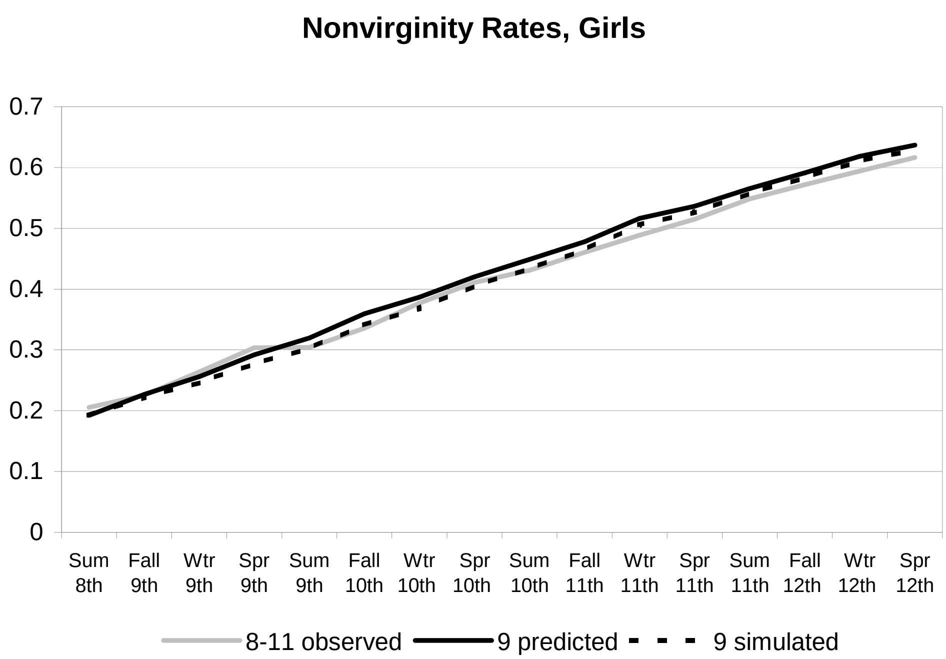


Figure 6b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is prediction from modified model.

Figure 7: Put Virgins and Nonvirgins in Different Schools at Initial Point in Time

Figure 7a.

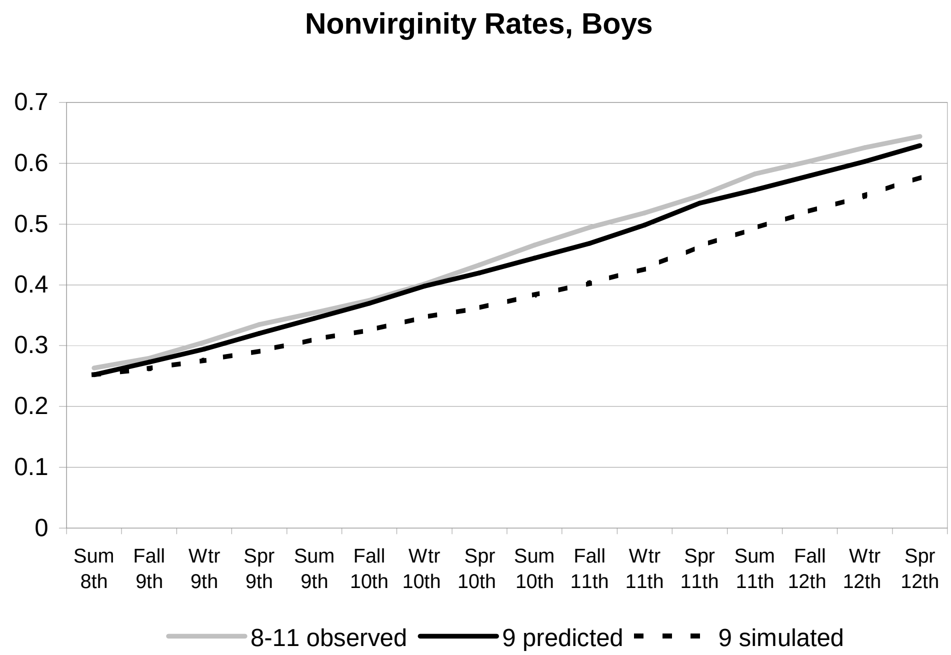
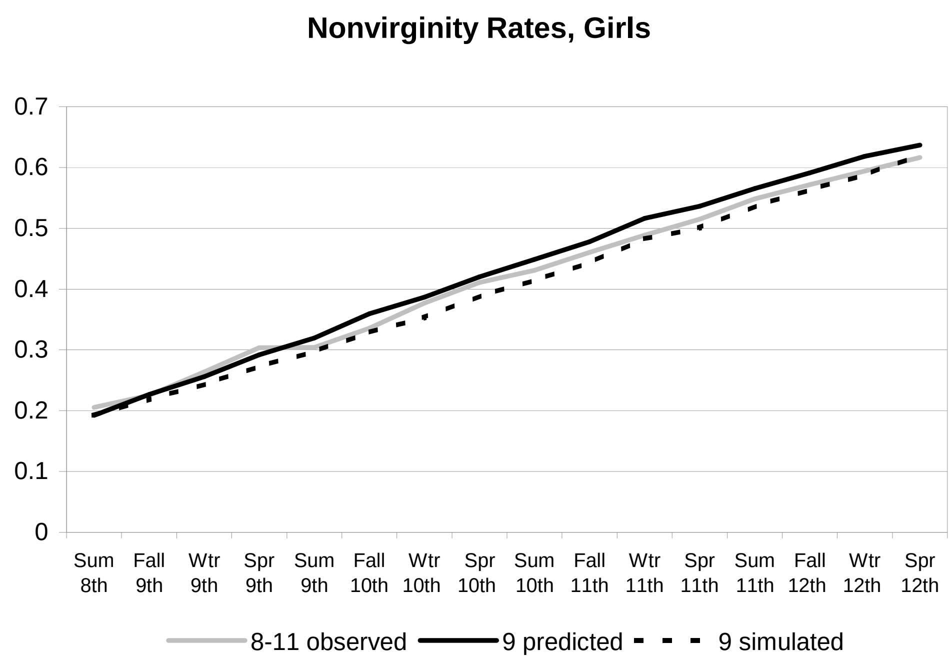


Figure 7b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is prediction from modified model.

Figure 8: "Low" Type Individuals Never Search (virginity pledge)

Figure 8a.

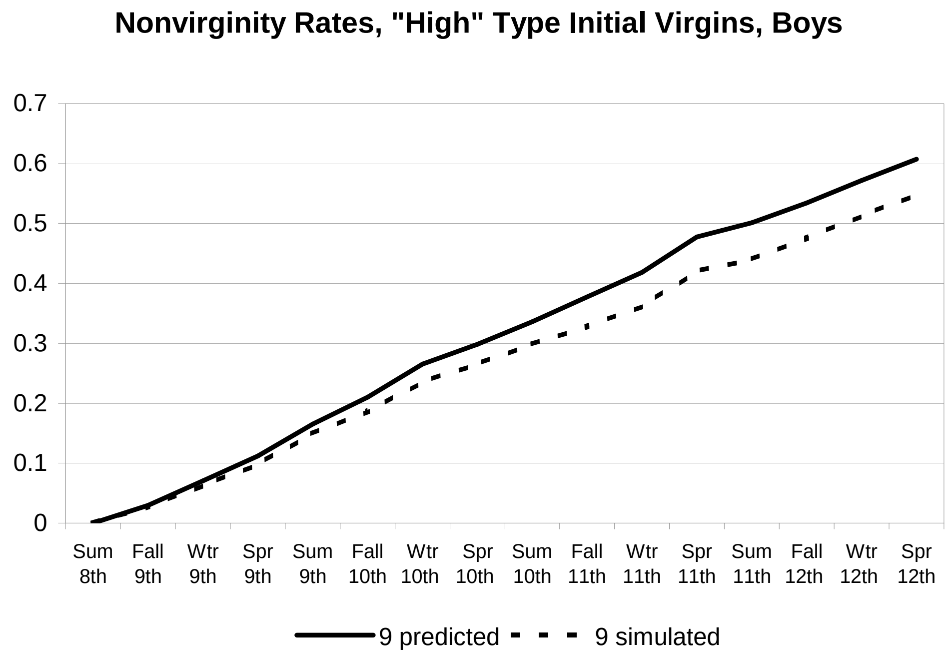
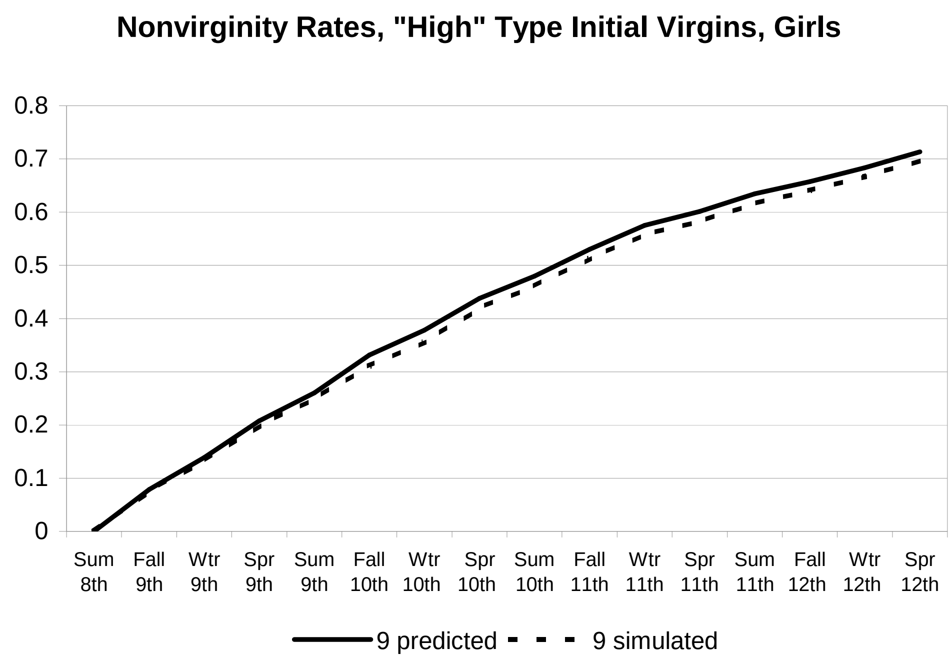


Figure 8b.



NOTES: "8-11 observed" is 8th-11th grade cohorts, combined; "9 predicted" is prediction for 9th grade cohort from estimated model; "9 simulated" is 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 in High 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: 2SLS Estimates of Effect of Social Interactions, Boys

	All Instruments Excluded				Contextual Effects Included				School-by-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**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	0.376**	0.564**	0.608**	0.517**	-0.331	1.160**	0.924**	0.787**	0.168	1.330**	0.744*	0.281
nonvirgin	(0.083)	(0.084)	(0.087)	(0.142)	(0.849)	(0.287)	(0.261)	(0.228)	(0.377)	(0.401)	(0.292)	(0.471)
Overid. test	8.87	13.26	19.80	3.41	0.00	4.08	3.76	0.91	6.38	5.43	6.33	2.41
p-value	0.26	0.07	0.01	0.85	0.95	0.04	0.05	0.34	0.50	0.61	0.50	0.93
PANEL B: Reference group is combined peer and supply groups												
Group share	0.519**	0.836**	0.812**	0.484**	0.746**	1.721**	1.706**	1.110*	0.962+	2.651**	1.801*	1.965*
nonvirgin	(0.093)	(0.095)	(0.129)	(0.120)	(0.235)	(0.294)	(0.513)	(0.546)	(0.518)	(0.579)	(0.825)	(0.999)
Overid. test	8.61	15.80	12.40	2.89	5.45	2.02	6.14	0.65	4.07	3.41	11.50	5.24
p-value	0.38	0.05	0.13	0.94	0.07	0.36	0.05	0.72	0.85	0.91	0.18	0.73

Robust standard errors in parentheses. Standard errors are clustered within reference groups, except for FE models.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 4: 2SLS Estimates of Effect of Social Interactions, Girls

	All Instruments Excluded				Contextual Effects Included				School-by-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**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	0.609**	0.690**	0.602**	0.660**	0.769**	0.805**	0.756**	0.763*	0.656	1.046+	0.956**	0.720*
nonvirgin	(0.092)	(0.086)	(0.085)	(0.100)	(0.165)	(0.150)	(0.220)	(0.353)	(0.505)	(0.543)	(0.371)	(0.320)
Overid. test	2.94	7.20	13.64	7.78	0.17	1.17	1.22	0.60	3.78	5.05	8.24	1.93
p-value	0.94	0.52	0.09	0.46	0.92	0.56	0.54	0.74	0.88	0.75	0.41	0.98
PANEL B: Reference group is combined peer and supply groups												
Group share	0.629**	0.820**	0.697**	0.751**	0.784**	0.689**	0.837**	0.909*	-0.404	0.135	1.455	1.431*
nonvirgin	(0.112)	(0.111)	(0.105)	(0.114)	(0.202)	(0.257)	(0.313)	(0.458)	(0.729)	(0.847)	(0.938)	(0.631)
Overid. test	8.44	11.93	14.59	7.84	4.10	0.95	0.69	0.32	7.73	7.95	10.55	3.82
p-value	0.39	0.15	0.07	0.45	0.13	0.62	0.71	0.85	0.46	0.44	0.23	0.87

Robust standard errors in parentheses. Standard errors are clustered within reference groups, except for FE models.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 5: Alternative Fixed Effects 2SLS Specifications, Boys

	Baseline FE Estimates				Limited Set of Variables				Limited Var. and Lagged Group Outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(5)	(6)	(7)	(8)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**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	0.168	1.330**	0.744*	0.281	-0.053	1.528**	0.751	0.779	-0.081	1.305**	0.896	1.004
nonvirgin	(0.377)	(0.401)	(0.292)	(0.471)	(0.447)	(0.484)	(0.502)	(0.743)	(0.448)	(0.463)	(0.578)	(0.993)
Overid. test	6.38	5.43	6.33	2.41	4.92	3.05	2.07	0.84	4.91	5.40	1.79	0.81
p-value	0.50	0.61	0.50	0.93	0.18	0.38	0.56	0.84	0.18	0.14	0.62	0.85
F-stat. on instr.	14.19	19.18	41.68	16.49	20.18	24.78	29.59	14.52	27.42	26.30	20.76	6.98
PANEL B: Reference group is combined peer and supply groups												
Group share	0.962+	2.651**	1.801*	1.965*	1.004	3.038**	6.222*	3.621+	1.393+	2.890**	3.228+	2.193
nonvirgin	(0.518)	(0.579)	(0.825)	(0.999)	(0.743)	(0.791)	(2.810)	(1.908)	(0.835)	(0.783)	(1.664)	(1.815)
Overid. test	4.07	3.41	11.50	5.24	1.84	1.15	0.17	0.48	0.75	2.46	2.36	3.29
p-value	0.85	0.91	0.18	0.73	0.61	0.76	0.98	0.92	0.86	0.48	0.50	0.35
F-stat. on instr.	40.06	91.51	68.08	18.55	49.32	68.50	8.32	12.33	55.85	103.93	14.34	9.11

Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 6: Alternative Fixed Effects 2SLS Specifications, Girls

	Baseline FE Estimates				Limited Set of Variables				Limited Var. and Lagged Group Outcomes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(5)	(6)	(7)	(8)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**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	0.656	1.046+	0.956**	0.720*	0.283	0.322	1.244+	0.597	0.092	0.294	1.710+	0.540
nonvirgin	(0.505)	(0.543)	(0.371)	(0.320)	(0.673)	(0.727)	(0.669)	(0.365)	(0.573)	(0.501)	(0.875)	(0.379)
Overid. test	3.78	5.05	8.24	1.93	1.64	2.44	0.78	0.66	1.72	2.29	0.16	1.40
p-value	0.88	0.75	0.41	0.98	0.65	0.49	0.85	0.88	0.63	0.51	0.98	0.70
F-stat. on instr.	5.34	14.10	20.17	24.07	8.69	12.73	19.88	49.76	16.10	23.84	8.97	43.45
PANEL B: Reference group is combined peer and supply groups												
Group share	-0.404	0.135	1.455	1.431*	-1.088	1.011	3.008*	1.018	-0.754	0.434	1.602	1.149
nonvirgin	(0.729)	(0.847)	(0.938)	(0.631)	(1.393)	(1.355)	(1.522)	(0.801)	(1.040)	(1.074)	(0.994)	(1.294)
Overid. test	7.73	7.95	10.55	3.82	4.38	3.66	0.96	1.91	4.52	4.28	2.37	2.76
p-value	0.46	0.44	0.23	0.87	0.22	0.30	0.81	0.59	0.21	0.23	0.50	0.43
F-stat. on instr.	35.76	26.49	14.04	18.27	18.36	19.61	11.00	26.17	38.27	28.07	21.49	11.37

Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 7: Structural Parameter Estimates

Parameter		Boys	Girls
Age			
	α	0.073	0.171
		(0.084)	(0.098)
Peer preference interaction			
	γ	0.222	0.147
		(0.086)	(0.094)
Arrival rate			
λ_0 :	9^{th} grade	-2.591	-2.394
		(0.301)	(0.253)
	10th grade	-2.884	-2.395
		(0.263)	(0.247)
	11th grade	-2.911	-2.403
		(0.277)	(0.241)
	12^{th} grade	-2.832	-2.296
		(0.298)	(0.248)
λ_{11} :	Same grade	0.468	0.085
		(0.368)	(0.280)
λ_{12} :	Down / 1 up (boys / girls)	0.339	0.103
		(0.251)	(0.159)
λ_{13} :	Up / 2 up (boys / girls)	0.253	0.187
		(0.182)	(0.156)

(continues next page)

Table 7. (continued)

Parameter		**Boys**	**Girls**
(continued)			
Type values			
	ω^L	-0.263	-0.290
		(0.090)	(0.099)
	ω^H	-0.108	-0.063
		(0.074)	(0.081)
Terminal values			
	Low type	-1.657	-0.085
		(1.087)	(1.083)
	High type	-0.198	0.272
		(1.208)	(3.872)
Type probabilities (π^H)			
	Constant	0.986	-1.797
		(1.223)	(1.129)
Y_0 :	9th grade	-1.075	0.248
		(2.973)	(1.592)
	10th grade	0.621	2.000
		(2.246)	(1.598)
	11th grade	-1.047	4.461
		(2.039)	(2.148)
	Black	3.535	0.209
		(2.955)	(0.419)
	Younger child	0.830	-0.108
		(0.685)	(0.341)
	Only child	2.351	2.074
		(2.151)	(1.063)
	Parent educ.	-2.822	-1.682
		(2.297)	(0.850)

Table 8: Probability of Search among Virgins, by Type, and Marginal Effects of Lagged Peer Nonvirginity Rates

	Boys					Girls			
	Search probability by type:			**Marg.**		**Search probability by type:**			**Marg.**
Grade	Low	High	Avg.	**Effects**		Low	High	Avg.	**Effects**
9th	0.160	0.484	0.324	0.404		0.162	0.694	0.412	0.233
10th	0.267	0.584	0.423	0.377		0.296	0.834	0.517	0.221
11th	0.389	0.644	0.510	0.274		0.424	0.921	0.594	0.150
12th	0.516	0.652	0.580	0.152		0.488	0.964	0.639	0.079

Note: "Avg." columns show the average search probability, which combines the search probability for the two types weighted by probability that an individual is of each type.

Table 9: Average Arrival Rates, and Marginal Effects of Search Behavior among the Opposite Gender

	Boys					**Girls**			
		Marginal effects of girls in:					**Marginal effects of boys in:**		
Grade	**Arrival Rate**	Same Grd.	Grd. Below	Grd. Above		**Arrival Rate**	Same Grd.	Grd. Above	Two Above
9^{th}	0.102	0.043	NA	0.023		0.104	0.008	0.010	0.017
10^{th}	0.098	0.041	0.030	0.022		0.106	0.008	0.010	0.018
11^{th}	0.105	0.044	0.032	0.024		0.094	0.007	0.009	NA
12^{th}	0.101	0.043	0.031	NA		0.097	0.007	NA	NA

Table 10: Type Distribution

10a) Probability of high type

Grade	Boys	Girls
9th	0.51	0.48
10th	0.50	0.41
11th	0.47	0.33
12th	0.49	0.33

10b) Effect of individual characteristics on probability of high type

Characteristic	Boys	Girls
Black race	0.46	0.03
Younger child	0.13	-0.02
Only child	0.33	0.34
Parental educ.	-0.46	-0.29

Table A-1: 2SLS Estimates for Being Sexually Experienced by Grade, using Same-Gender Peer Group, Boys

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group												
Peer group	0.376**	0.564**	0.608**	0.517**	-0.331	1.160**	0.924**	0.787**	0.168	1.330**	0.744*	0.281
	(0.083)	(0.084)	(0.087)	(0.142)	(0.849)	(0.287)	(0.261)	(0.228)	(0.377)	(0.401)	(0.292)	(0.471)
Individual characteristics												
Hispanic	0.153**	0.179**	0.112**	0.061	0.177**	0.213**	0.141**	0.077+	0.171**	0.200**	0.143**	0.074
	(0.040)	(0.040)	(0.038)	(0.038)	(0.043)	(0.047)	(0.044)	(0.046)	(0.048)	(0.048)	(0.044)	(0.046)
Black	0.240**	0.202**	0.164**	0.105**	0.264**	0.206**	0.189**	0.122**	0.252**	0.213**	0.188**	0.127**
	(0.040)	(0.035)	(0.035)	(0.036)	(0.043)	(0.043)	(0.040)	(0.045)	(0.043)	(0.043)	(0.040)	(0.040)
Younger child	0.064*	0.056*	0.033	-0.003	0.056*	0.047+	0.028	-0.002	0.062*	0.047	0.026	-0.003
	(0.026)	(0.025)	(0.029)	(0.031)	(0.027)	(0.026)	(0.030)	(0.031)	(0.027)	(0.029)	(0.029)	(0.029)
Only child	0.146**	0.093**	0.067*	0.099**	0.148**	0.087**	0.056	0.090*	0.148**	0.087*	0.064+	0.090**
	(0.032)	(0.029)	(0.034)	(0.037)	(0.033)	(0.031)	(0.037)	(0.038)	(0.035)	(0.036)	(0.036)	(0.035)
Mother married	-0.111**	-0.076**	-0.062*	0.020	-0.109**	-0.083**	-0.072*	0.018	-0.120**	-0.092**	-0.074*	0.022
	(0.031)	(0.026)	(0.030)	(0.029)	(0.032)	(0.029)	(0.032)	(0.031)	(0.030)	(0.031)	(0.029)	(0.028)
Foreign-born	-0.101*	-0.133**	-0.076*	-0.053	-0.092*	-0.106**	-0.047	-0.046	-0.072	-0.109*	-0.040	-0.049
parent	(0.044)	(0.033)	(0.031)	(0.038)	(0.047)	(0.037)	(0.037)	(0.043)	(0.045)	(0.043)	(0.041)	(0.048)
Upper income	-0.020	-0.027	-0.003	-0.001	-0.020	-0.034	-0.008	-0.003	-0.016	-0.035	-0.005	0.000
	(0.028)	(0.031)	(0.031)	(0.033)	(0.029)	(0.034)	(0.033)	(0.034)	(0.031)	(0.034)	(0.033)	(0.033)
Parental education	-0.122**	-0.100**	-0.092**	-0.077*	-0.130**	-0.092**	-0.080**	-0.075*	-0.106**	-0.105**	-0.081**	-0.080*
(4 years college)	(0.025)	(0.028)	(0.030)	(0.034)	(0.032)	(0.029)	(0.029)	(0.035)	(0.027)	(0.030)	(0.031)	(0.031)

(continues next page)

Table A-1. (continued)

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Contextual effects of mean peer characteristics												
Hispanic	---	---	---	---	0.094	-0.124	-0.148+	-0.133	---	---	---	---
					(0.156)	(0.085)	(0.079)	(0.112)				
Black	---	---	---	---	0.207	-0.221+	-0.143	-0.091	---	---	---	---
					(0.299)	(0.117)	(0.107)	(0.063)				
Mother married	---	---	---	---	-0.014	0.071	0.058	-0.027	---	---	---	---
					(0.135)	(0.062)	(0.054)	(0.063)				
Foreign-born parent	---	---	---	---	-0.393	0.052	0.047	0.133	---	---	---	---
					(0.293)	(0.107)	(0.101)	(0.131)				
Upper income	---	---	---	---	-0.032	0.119+	0.104	0.054	---	---	---	---
					(0.127)	(0.065)	(0.067)	(0.062)				
Parental education (4 years college)	---	---	---	---	-0.291	0.172	-0.017	0.015	---	---	---	---
					(0.301)	(0.127)	(0.128)	(0.073)				
Overidentification test	8.87	13.26	19.80	3.41	0.00	4.08	3.76	0.91	6.38	5.43	6.33	2.41
p-value	0.26	0.07	0.01	0.85	0.95	0.04	0.05	0.34	0.50	0.61	0.50	0.93
F-stat., first-stage instr.	35.33	24.13	18.15	6.68	1.27	3.94	2.95	6.29	14.19	19.18	41.68	16.49
p-value	< 0.001	< 0.001	< 0.001	< 0.001	0.284	0.021	0.055	0.002	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2241	2713	2873	2908	2241	2713	2873	2908	2241	2713	2873	2908

Models without fixed effects also include dummies for urban school districts and districts where the ninth grade is separated from the rest of high school.
Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-2: 2SLS Estimates for Being Sexually Experienced by Grade, using Same-Gender Peer Group, Girls

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group												
Peer group	0.609**	0.690**	0.602**	0.660**	0.769**	0.805**	0.756**	0.763*	0.656	1.046+	0.956**	0.720*
	(0.092)	(0.086)	(0.085)	(0.100)	(0.165)	(0.150)	(0.220)	(0.353)	(0.505)	(0.543)	(0.371)	(0.320)
Individual characteristics												
Early menarche	0.065*	0.089**	0.082**	0.097**	0.061*	0.088**	0.081**	0.095**	0.065*	0.085**	0.078**	0.094**
	(0.028)	(0.024)	(0.024)	(0.026)	(0.029)	(0.024)	(0.025)	(0.026)	(0.027)	(0.030)	(0.029)	(0.026)
Hispanic	0.068*	0.012	-0.023	-0.059	0.087*	0.030	-0.008	-0.039	0.097*	0.049	-0.003	-0.034
	(0.033)	(0.040)	(0.040)	(0.039)	(0.038)	(0.046)	(0.047)	(0.047)	(0.046)	(0.047)	(0.045)	(0.041)
Black	0.069*	0.051	0.043	0.023	0.082*	0.043	0.076+	0.055	0.082+	0.050	0.094*	0.064+
	(0.028)	(0.034)	(0.029)	(0.025)	(0.033)	(0.047)	(0.041)	(0.038)	(0.045)	(0.040)	(0.039)	(0.037)
Younger child	-0.031	-0.011	0.004	0.041	-0.034	-0.011	0.005	0.043	-0.023	-0.012	-0.001	0.037
	(0.029)	(0.027)	(0.029)	(0.032)	(0.029)	(0.028)	(0.030)	(0.032)	(0.027)	(0.030)	(0.030)	(0.027)
Only child	0.055	0.083*	0.084*	0.115**	0.052	0.079*	0.083*	0.115**	0.052	0.073+	0.081*	0.099**
	(0.039)	(0.036)	(0.034)	(0.032)	(0.040)	(0.038)	(0.034)	(0.033)	(0.036)	(0.038)	(0.037)	(0.031)
Mother married	-0.050+	-0.039	-0.056*	-0.063**	-0.052+	-0.040	-0.058*	-0.061**	-0.052+	-0.043	-0.052+	-0.060*
	(0.028)	(0.026)	(0.026)	(0.024)	(0.029)	(0.027)	(0.027)	(0.024)	(0.029)	(0.030)	(0.029)	(0.026)
Foreign-born parent	-0.082+	-0.038	-0.093*	-0.085*	-0.077	-0.033	-0.087+	-0.085*	-0.070	-0.029	-0.082+	-0.085*
	(0.043)	(0.043)	(0.043)	(0.035)	(0.047)	(0.046)	(0.045)	(0.039)	(0.045)	(0.051)	(0.049)	(0.041)
Upper income	-0.013	0.010	-0.003	0.011	-0.012	0.008	-0.012	0.005	-0.009	0.008	-0.016	0.010
	(0.025)	(0.029)	(0.030)	(0.039)	(0.026)	(0.030)	(0.032)	(0.040)	(0.030)	(0.033)	(0.034)	(0.032)
Parental education (4 years college)	-0.100**	-0.094**	-0.125**	-0.121**	-0.099**	-0.099**	-0.123**	-0.119**	-0.097**	-0.100**	-0.124**	-0.122**
	(0.025)	(0.026)	(0.034)	(0.029)	(0.025)	(0.026)	(0.035)	(0.032)	(0.026)	(0.031)	(0.033)	(0.031)

(continues next page)

Table A-2. (continued)

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Contextual effects of mean peer characteristics												
Hispanic	---	---	---	---	-0.045	-0.066	-0.069	-0.110	---	---	---	---
					(0.055)	(0.085)	(0.088)	(0.111)				
Black	---	---	---	---	-0.052	0.015	-0.063	-0.082+	---	---	---	---
					(0.058)	(0.055)	(0.049)	(0.048)				
Mother married	---	---	---	---	0.055	-0.020	-0.049	-0.005	---	---	---	---
					(0.048)	(0.047)	(0.053)	(0.076)				
Foreign-born parent	---	---	---	---	-0.013	0.066	0.131	0.133+	---	---	---	---
					(0.080)	(0.079)	(0.095)	(0.080)				
Upper income	---	---	---	---	-0.015	0.084	0.215**	-0.006	---	---	---	---
					(0.046)	(0.061)	(0.052)	(0.085)				
Parental education (4 years college)	---	---	---	---	0.033	0.066	-0.054	0.048	---	---	---	---
					(0.072)	(0.078)	(0.096)	(0.157)				
Overidentification test	2.94	7.20	13.64	7.78	0.17	1.17	1.22	0.60	3.78	5.05	8.24	1.93
p-value	0.94	0.52	0.09	0.46	0.92	0.56	0.54	0.74	0.88	0.75	0.41	0.98
F-stat., first-stage instr.	14.54	10.44	13.49	8.55	3.73	5.52	1.84	1.31	5.34	14.10	20.17	24.07
p-value	< 0.001	< 0.001	< 0.001	< 0.001	0.0127	0.0013	0.1418	0.2731	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2381	2849	2963	3141	2381	2849	2963	3141	2381	2849	2963	3141

Models without fixed effects also include dummies for urban school districts and districts where the ninth grade is separated from the rest of high school.
Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-3: 2SLS Estimates for Being Sexually Experienced by Grade, using Peer and Supply Groups, Boys

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group												
Peer and supply groups combined	0.519**	0.836**	0.812**	0.484**	0.746**	1.721**	1.706**	1.110*	0.962+	2.651**	1.801*	1.965*
	(0.093)	(0.095)	(0.129)	(0.120)	(0.235)	(0.294)	(0.513)	(0.546)	(0.518)	(0.579)	(0.825)	(0.999)
Individual characteristics												
Hispanic	0.166**	0.203**	0.152**	0.077*	0.189**	0.210**	0.149**	0.079+	0.176**	0.197**	0.138**	0.088+
	(0.035)	(0.041)	(0.035)	(0.039)	(0.041)	(0.054)	(0.043)	(0.046)	(0.049)	(0.045)	(0.044)	(0.047)
Black	0.231**	0.196**	0.184**	0.114**	0.249**	0.195**	0.188**	0.122*	0.260**	0.203**	0.184**	0.128**
	(0.040)	(0.034)	(0.037)	(0.038)	(0.048)	(0.042)	(0.048)	(0.048)	(0.043)	(0.041)	(0.039)	(0.040)
Younger child	0.063*	0.057*	0.042	-0.002	0.064*	0.058*	0.042	-0.001	0.064*	0.047+	0.032	-0.000
	(0.027)	(0.025)	(0.026)	(0.030)	(0.027)	(0.025)	(0.027)	(0.030)	(0.027)	(0.028)	(0.028)	(0.029)
Only child	0.146**	0.092**	0.069*	0.100**	0.149**	0.093**	0.067*	0.091**	0.153**	0.085*	0.070*	0.087*
	(0.030)	(0.028)	(0.032)	(0.033)	(0.030)	(0.029)	(0.034)	(0.034)	(0.035)	(0.035)	(0.035)	(0.036)
Mother married	-0.111**	-0.075*	-0.065*	0.016	-0.113**	-0.073*	-0.062*	0.019	-0.120**	-0.075*	-0.075**	0.025
	(0.035)	(0.030)	(0.032)	(0.028)	(0.035)	(0.030)	(0.031)	(0.027)	(0.030)	(0.030)	(0.029)	(0.028)
Foreign-born parent	-0.090*	-0.113**	-0.051	-0.037	-0.070	-0.088*	-0.046	-0.049	-0.067	-0.107**	-0.056	-0.043
	(0.042)	(0.033)	(0.035)	(0.041)	(0.044)	(0.035)	(0.042)	(0.043)	(0.045)	(0.040)	(0.043)	(0.048)
Upper income	-0.016	-0.022	0.006	0.001	-0.018	-0.040	-0.007	-0.003	-0.020	-0.044	-0.000	-0.003
	(0.028)	(0.026)	(0.032)	(0.033)	(0.029)	(0.029)	(0.036)	(0.034)	(0.031)	(0.033)	(0.033)	(0.034)
Parental education (4 years college)	-0.114**	-0.085**	-0.080**	-0.075**	-0.109**	-0.084**	-0.073*	-0.070*	-0.103**	-0.106**	-0.084**	-0.081*
	(0.022)	(0.030)	(0.028)	(0.028)	(0.024)	(0.030)	(0.029)	(0.028)	(0.027)	(0.029)	(0.031)	(0.032)

(continues next page)

Table A-3. (continued)

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Contextual effects of mean peer/supply characteristics												
Hispanic	---	---	---	---	0.068	0.278+	0.343*	0.178	---	---	---	---
					(0.118)	(0.149)	(0.168)	(0.177)				
Black	---	---	---	---	-0.112	-0.105	-0.045	-0.039	---	---	---	---
					(0.099)	(0.091)	(0.096)	(0.069)				
Mother married	---	---	---	---	0.003	0.341**	0.355*	0.120	---	---	---	---
					(0.121)	(0.115)	(0.159)	(0.126)				
Foreign-born	---	---	---	---	-0.169	-0.077	-0.036	0.098	---	---	---	---
parent					(0.145)	(0.193)	(0.168)	(0.146)				
Upper income	---	---	---	---	-0.055	-0.065	0.026	-0.073	---	---	---	---
					(0.096)	(0.138)	(0.146)	(0.116)				
Parental education	---	---	---	---	0.177	0.570**	0.425	0.308	---	---	---	---
(4 years college)					(0.170)	(0.207)	(0.350)	(0.314)				
Overidentification test	8.61	15.80	12.40	2.89	5.45	2.02	6.14	0.65	4.07	3.41	11.50	5.24
p-value	0.38	0.05	0.13	0.94	0.07	0.36	0.05	0.72	0.85	0.91	0.18	0.73
F-stat., first-stage instr.	20.91	28.39	20.06	10.72	4.72	8.02	3.41	2.80	40.06	91.51	68.08	18.55
p-value	< 0.001	< 0.001	< 0.001	< 0.001	0.005	< 0.001	0.022	0.045	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2241	2713	2873	2908	2241	2713	2873	2908	2241	2713	2873	2908

Models without fixed effects also include dummies for urban school districts and districts where the ninth grade is separated from the rest of high school.
Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-4: 2SLS Estimates for Being Sexually Experienced by Grade, using Peer and Supply Groups, Girls

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Nonvirginity rate among reference group												
Peer and supply	0.629**	0.820**	0.697**	0.751**	0.784**	0.689**	0.837**	0.909*	-0.404	0.135	1.455	1.431*
groups combined	(0.112)	(0.111)	(0.105)	(0.114)	(0.202)	(0.257)	(0.313)	(0.458)	(0.729)	(0.847)	(0.938)	(0.631)
Individual characteristics												
Early menarche	0.064*	0.083**	0.082**	0.096**	0.062*	0.086**	0.083**	0.095**	0.067*	0.078**	0.084**	0.094**
	(0.025)	(0.023)	(0.022)	(0.025)	(0.025)	(0.023)	(0.022)	(0.026)	(0.026)	(0.028)	(0.028)	(0.026)
Hispanic	0.061+	-0.023	-0.052	-0.076*	0.101**	0.020	-0.018	-0.043	0.095*	0.046	-0.010	-0.035
	(0.033)	(0.037)	(0.040)	(0.034)	(0.037)	(0.040)	(0.045)	(0.042)	(0.045)	(0.044)	(0.043)	(0.042)
Black	0.058*	0.025	0.020	0.007	0.087**	0.041	0.081*	0.054	0.076+	0.054	0.096*	0.064+
	(0.028)	(0.036)	(0.031)	(0.029)	(0.033)	(0.044)	(0.040)	(0.040)	(0.044)	(0.038)	(0.038)	(0.038)
Younger child	-0.032	-0.007	0.003	0.039	-0.035	-0.004	0.005	0.041	-0.019	-0.002	-0.000	0.038
	(0.028)	(0.025)	(0.029)	(0.027)	(0.027)	(0.025)	(0.028)	(0.027)	(0.027)	(0.027)	(0.029)	(0.027)
Only child	0.050	0.091*	0.089**	0.112**	0.046	0.094*	0.087**	0.114**	0.050	0.082*	0.085*	0.101**
	(0.037)	(0.037)	(0.033)	(0.027)	(0.038)	(0.038)	(0.033)	(0.028)	(0.035)	(0.035)	(0.036)	(0.031)
Mother married	-0.046+	-0.036	-0.055*	-0.067**	-0.049+	-0.039	-0.055*	-0.064**	-0.049+	-0.043	-0.050+	-0.060*
	(0.026)	(0.028)	(0.027)	(0.020)	(0.026)	(0.028)	(0.027)	(0.020)	(0.029)	(0.029)	(0.028)	(0.026)
Foreign-born	-0.077+	-0.035	-0.113**	-0.094**	-0.056	-0.013	-0.088*	-0.087*	-0.067	-0.019	-0.081	-0.092*
parent	(0.039)	(0.040)	(0.042)	(0.032)	(0.041)	(0.044)	(0.043)	(0.036)	(0.042)	(0.047)	(0.049)	(0.042)
Upper income	-0.005	0.015	0.000	0.012	0.001	0.010	-0.015	0.003	-0.008	0.005	-0.021	0.006
	(0.027)	(0.031)	(0.033)	(0.036)	(0.029)	(0.031)	(0.036)	(0.037)	(0.029)	(0.031)	(0.033)	(0.032)
Parental education	-0.099**	-0.099**	-0.125**	-0.117**	-0.098**	-0.097**	-0.120**	-0.115**	-0.098**	-0.092**	-0.118**	-0.122**
(4 years college)	(0.025)	(0.023)	(0.036)	(0.033)	(0.025)	(0.024)	(0.037)	(0.035)	(0.026)	(0.030)	(0.033)	(0.032)

(continues next page)

Table A-4. (continued)

	All Instruments Excluded				Contextual Effects				School-Grade Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grade 9	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**	**Grade 9**	**Grade 10**	**Grade 11**	**Grade 12**
Contextual effects of mean peer/supply characteristics												
Hispanic	---	---	---	---	-0.132	-0.210+	-0.175	-0.134	---	---	---	---
					(0.102)	(0.124)	(0.111)	(0.179)				
Black	---	---	---	---	-0.127	0.004	-0.169*	-0.142*	---	---	---	---
					(0.088)	(0.101)	(0.073)	(0.066)				
Mother married	---	---	---	---	-0.002	-0.067	-0.123	-0.073	---	---	---	---
					(0.094)	(0.108)	(0.081)	(0.086)				
Foreign-born	---	---	---	---	-0.041	0.010	0.059	0.114	---	---	---	---
parent					(0.158)	(0.150)	(0.106)	(0.109)				
Upper income	---	---	---	---	-0.046	0.245*	0.256**	0.038	---	---	---	---
					(0.099)	(0.123)	(0.088)	(0.123)				
Parental education	---	---	---	---	0.055	-0.270+	-0.139	0.063	---	---	---	---
(4 years college)					(0.147)	(0.162)	(0.167)	(0.252)				
Overidentification test	8.44	11.93	14.59	7.84	4.10	0.95	0.69	0.32	7.73	7.95	10.55	3.82
p-value	0.39	0.15	0.07	0.45	0.13	0.62	0.71	0.85	0.46	0.44	0.23	0.87
F-stat., first-stage instr.	34.12	24.95	10.18	10.1	8.90	10.90	5.97	2.58	35.76	26.49	14.04	18.27
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.001	0.059	< 0.001	< 0.001	< 0.001	< 0.001
Observations	2381	2849	2963	3141	2381	2849	2963	3141	2381	2849	2963	3141

Models without fixed effects also include dummies for urban school districts and districts where the ninth grade is separated from the rest of high school.
Robust standard errors in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Table A-5: Fixed Effects 2SLS with Separate Instruments by Gender

	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**
Reference group is combined peer and supply groups								
Group share	0.944*	2.797**	1.116	1.092+	0.147	0.719	1.101+	0.824+
nonvirgin	(0.412)	(0.554)	(0.739)	(0.569)	(0.672)	(0.612)	(0.579)	(0.462)
Overid. test	9.59	6.10	21.45	6.11	13.42	11.50	20.21	10.21
p-value	0.89	0.99	0.16	0.99	0.64	0.78	0.21	0.86
F-stat. on instr.	70.44	78.58	50.22	34.83	23.85	35.01	27.82	24.99