Endogenous Sibling Effects in Adolescent Substance Use

Mithuna Srinivasan *

November, 2010

*The Ohio State University, Department of Economics, 309 Arps Hall, 1945 N High Street, Columbus OH 43210; email: srinivasan.66@osu.edu. I thank Bruce Weinberg, Audrey Light and Stephen Cosslett for comments and suggestions from the initial stages of this research.

Abstract

Longitudinal data shows that the likelihood of a child smoking more than triples if the child has an older sibling who also smoked. The current study unpacks this striking correlation, asking whether it is causal or a result of correlated unobservables such as parental investments that are endogenous to the child's behavior. In particular, I examine whether smoking or drinking by an older sibling influences the corresponding behavior of a younger sibling. To control for endogeneity in the older child's outcome, I use differences in smoking and drinking by gender and age among older siblings as instrumental variables. While previous studies have established gender differences in teen substance use, I find that these differences vary across age. For example, gender differences in drinking are small at younger ages, but males drink consistently more than females from mid to late adolescence. The instruments are plausibly exogenous of unobservables such as parental substance use, and will control for parental investments under the assumption that the older sibling's age and gender do not directly affect resources invested in the younger child. I empirically investigate this assumption using data on measures of parental investments, and find no evidence of a correlation between the instruments and younger sibling investments. The results point toward significant and positive sibling effects for smoking as well as drinking. These findings indicate the presence of opportunities for resource constrained parents to invest efficiently in favor of their firstborn, to reduce "bad" behavior. Positive sibling effects imply that curtailing the older child's behavior in this way can have spillover effects on the younger sibling, leading to greater payoffs to parents in overall child quality. Within-family social multipliers may also serve to amplify the effects of public policies aimed at curtailing smoking and drinking.

1 Introduction

The proportions of 8th, 10th, and 12th graders who admitted drinking an alcoholic beverage in a 30-day period are about 15%, 30%, and 44% respectively. The statistics for smoking are just as sobering.[1] Social scientists have stressed the importance of familial interactions in influencing adolescent substance use. Research has shown that although parental smoking and drinking are strong risk factors, peer use and use by siblings may be even more influential (Fagan and Najman (2005), Windle (2000)). Data from the National Longitudinal Survey of Youth 1997 (NLSY97) shows that the likelihood of smoking in the past month by a younger child increases significantly from 12% to about 38% if he has an older sibling who also smoked in the same period of time.[2] These striking correlations suggest possible causal influences from older to younger siblings in adolescent substance use.[3] The current study unpacks this correlation, investigating whether smoking and drinking by an older sibling has a direct effect on the corresponding behavior of his or her younger sibling, by using instrumental variable (IV) estimation.[4]

To estimate the causal effect of an older sibling's smoking and drinking on the corresponding

[1] See "Monitoring the Future" reports.

[2] See Table 3. Other studies that have examined whether siblings' use of such substances are related also report strong correlations (Duncan et al (2001), Boyle et al (2001)).

[3] Sociological and psychological theories generally emphasize that age is an important determinant of interpersonal relationships as it is associated with power and influence (Parsons and Bales, 1955). Older individuals have more experience and knowledge in areas of interest to younger individuals and so in general, older siblings are often looked at as influential sources of guidance, support and knowledge for younger siblings owing to the hierarchial nature of their relationship (Jenkins, Barber, and Eccles, 1997). There is also some evidence to show that younger siblings are more reliant on sibling interactions, depict greater admiration for their older brothers and sisters and consequently imitate them more (Buhrmester (1992), Pepler et al (1981)). Thus, researchers have posited that there exist stronger grounds for looking at older siblings as likely referents for their younger counterparts, rather than the other way around.

[4] The sibling influences estimated in this study are related to the larger literature on peer effects and social interactions (eg., Clark and Loheac (2006), Gaviria and Raphael (2001), Powell et al (2005), Evans, Oats and Schwab (1992)). Most studies dealing with school going peers specify the reference group in a rather ad hoc manner, usually conditional on the availability of data. Defining the peer group clearly as consisting of the older child highlights one of the advantages of using family peers to study interaction effects. Despite the empirically significant correlations between siblings in risky behavior, it is surprising that there exists very little quantitative evidence till date on sibling effects in the economics literature, and almost all studies have been limited to school based peers as a source of behavioral influences. An interesting paper in economics pertaining to siblings is Argys et al (2006) find that middle borns and last borns are most likely to use illicit substances and be sexually active as compared to their firstborn counterparts which suggests the importance of older siblings in risky behavior. One of the few published economics studies dealing with direct sibling influences is Oettinger (2000) who used a sample of pairs from the NLSY to study sibling effects in the probability of high school graduation. He found that increases in the older child's achievement raises younger siblings achievement.

behavior of a younger child, the older sibling's reported substance use must be purged of its correlation with the error term in the younger sibling's regression equation. Such a correlation can arise because of unobserved differences across families in the shared family environment of siblings. I control for a wide range of observable family background characteristics such as maternal education, family structure, family income, race and maternal age which can account for some part of the unobserved heterogeneity across families. Variability across groups in family background characteristics can also be generated by factors such as parental substance use, that influence an adolescent's unobservable "taste" for smoking or drinking. Across-family heterogeneity can also ensue from other, non-family factors that are common between siblings. School quality is one such determinant. Siblings from low quality schools with poor teachers, bad friends and neighborhoods may display a greater likelihood for substance use as compared to children from better quality institutions with clear-cut school policies designed to discourage such behavior. Data limitations prevent the adequate measurement of all such influences on adolescent substance use. These unobservables will reflect in a positive correlation between the substance use of siblings, which can be wrongly interpreted as a causal sibling effect.[5]

Another source of endogeneity in the older sibling's behavior can be attributed to parental investments made in children. Parents can choose to invest in many ways through financial transfers, non-monetary measures such as supervision and monitoring or investments in children's cultural capital. Families can differ in the quantity and quality of investments they make in their children. Families in which both children engage in smoking or drinking may be the ones that for unobserved reasons do not invest much in their offspring, while "better" behaved siblings may be a reflection of better and more consistent parenting.[6] However, accounting for investments is a challenging

[5]This is related to the selection bias issue in the peer effects literature, where spurious effects arise when youths belonging to the same reference group behave similarly since they sort across neighborhoods, schools or friends circles on the basis of a common set of unobserved characteristics (see Manski (1993, 2000)). Although siblings are born into the same family and do not select one another, correlated effects ensuing out of unobserved common influences such as parental substance use and school quality among others, can still induce bias in the estimates of sibling influences.

[6]Developmental researchers have generally found that adolescents from homes with high levels of parental investment such as monitoring are usually less likely to engage in the consumption of alcohol and drugs. For example, Lamborn and others (1991) suggest that authoritative parenting might help to curtail adolescent experimentation with illicit substances such as drugs and alcohol through the emphasis on behavioral accountability together with parental monitoring of offspring behavior and whereabouts. Other authors who support this assertion are Cohen and Rice (1997) and Steinberg et al (1994).

task given their multi-faceted nature, so data that captures most dimensions is seldom available. If the older sibling's smoking or drinking reflects these omitted investments, a positive coefficient on the older child's behavior will result even if there are no direct effects of the older sibling's reported substance use on the substance use of the younger.

I use IVs to address the endogeneity in the older sibling's behavior, using the age and gender of the older child as my key source of identification in his or her outcome. To illustrate, Figure 1 plots the number of days on which older siblings aged 13 to 18 years from two children families drank alcohol in the month prior to the survey, by gender. In the early teenage years, boys and girls have similar average rates of alcohol consumption, but males drink consistently more than females in mid to late adolescence.[7] Peer pressure is usually given heightened importance in discussions of adolescent substance use and the gendered patterns emerging from Figure 1 may be consistent with gender differences in the susceptibility to peer pressure.[8]

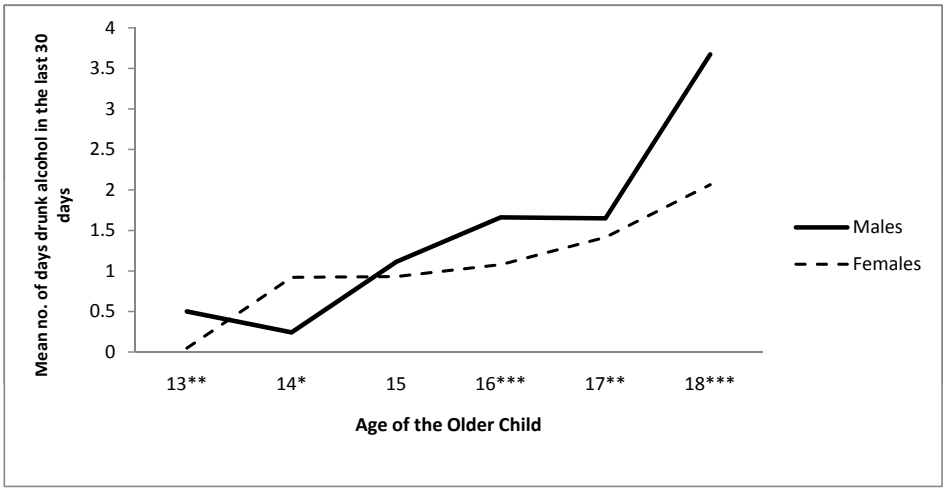
Although there are numerous papers in sociology and psychology that deal with the estimation of direct sibling influences in substance use, the main limitation of these studies is that most of them have not used appropriate techniques that take care of shared unobservables to address the issue of endogenous sibling behavior. In contrast, my instruments namely, the older sibling's age and gender, allow me to control for the effects of most shared unobservables driving the younger sibling's behavior such as price, parental substance use and school quality.[9] Since models of

[7]Other studies have shown that gender differences exist among adolescents for substance use outcomes such as drinking of alcohol, and they typically assert that males show higher levels of current alcohol consumption and binge drinking as compared to females (Barnes et al (2002), Johnston, O'Malley and Bachman (2000)). According to the 2005 survey of the National Survey on Drug Use and Health, more males than females aged 12 to 20 reported current alcohol use as well as heavy drinking. Recent results from the survey "Monitoring the Future" indicate that gender differences in substance use do emerge at older ages among adolescents, with males drinking more heavily than females.

[8]Males usually face pressures to establish a masculine identity and increase their status, reputation and physical attractiveness as "men" prior to entering the adult world. Since boys normally mature later than girls, they may become conscious of their physicality more during periods of mid to late adolescence and consequently have lower resistance to peer pressure in these ages.

[9]One might argue that gender of children can be endogenously determined through sex selective abortion. However, gender selective procedures are more prevalent among developing countries and are not much of an issue for industrialized economies such as the US. It has been found that parents in the US typically prefer a gender mix among their children. In fact, since the legalization of abortion, the birth ratio of girls has increased relative to boys. Other procedures such as sperm sorting and IVF have not only been introduced recently, but are also expensive and not very popular. In general, even if some parents display a gender bias in favor of male children, this does not take the extreme form of "missing girls" in the US as in some Asian countries (Dahl and Moretti, 2004).

Figure 1 : Gender Differences in Drinking Among Older Children in the NLSY97 : Unconditional Data



*, ** and *** denote significance of age-wise gender differences at 1%, 5% and 10% levels respectively.

intrafamily allocation suggest that parental preferences for investing in their offspring can be driven by child characteristics (Behrman, Pollak and Taubman, 1982), it is reasonable to expect that the younger sibling's personal demographics affect investments made in him or her. However, I assume that the *older* sibling's age and gender have no direct effect on investments made in the younger child, through parental allocation decisions, *all else equal*.[10] It is reasonable to believe that after controlling for the younger child's personal characteristics, the older sibling's characteristics should not have any additional effect. However, if parents face time or income constraints, it is possible that gender or birth order preferences driven by the older sibling's characteristics can affect investments made in the younger child. Since I control for the main effects of family income, there

[10]It should be noted that this assumption does not imply that the sibship sex composition has no effect on parental investments. We can expect the older child's gender to affect younger child investments within particular types of families defined by the gender composition of the sibling pair. For example, younger girls with older brothers can receive lower investments as compared to younger girls with older sisters if parents harbor a pro-male bias, even in the absence of time or budget constraints. This type of a comparison entails specifying gender composition dummies, for eg., boy-boy pairs, boy-girl pairs, etc. However, there is no apparent reason to suspect why investments in younger children in *all* sibship types will vary systematically with their older sibling's gender. With respect to an age preference, we might expect families to invest differently in their younger children depending on how old the older child is. For example, parents with firstborn children aged 17 might invest differently in younger siblings compared to families where the first child is aged 14. However, such differences may be a more severe problem in sibships with wide age gaps. In my sample, since the average age difference between siblings is two years, any age effect on investments will already be captured by the younger sibling's age.

would presumably be no across-family variability in budget constraints facing mothers and fathers. Across-family differences may however be produced by differences between groups in the "costs" of spending time with children.[11] It should be noted that I control for a wide range of observables that place families in different structural positions such as family structure and socioeconomic status. Assuming that these variables are a reasonable approximation of parental labor supply, even if there are remaining unobservable between-family differences in work or time constraints, I argue that they may be small.[12] In any case, rather than just invoking the identification assumption, I am able to actually investigate its empirical validity with the investment data available in the NLSY97. The data support this identification assumption remarkably well. I find no evidence for a significant correlation between the instruments and younger child investments. Although these data do not capture all dimensions of parental investments, the IV estimates are valid under the assumption that other unobserved components have similar distributional properties.

The main finding of the paper is that there exist significant and positive sibling effects for both types of substance use, measured by smoking and drinking. This is qualitatively consistent with most other studies of sibling influences in adolescent risky behavior in psychology and sociology (see Haurin and Mott (1990), Fagan and Najman (2005), Windle (2000)). Positive influences imply that as the older child smokes or drinks more, it leads to a corresponding increase in such behavior of the younger child. There are a variety of interpretations for these findings. Increased substance use by the older sibling can affect the marginal benefit of the same action for the younger child, through role modeling by the older sibling (or imitation by the younger child) and/or indirectly through the selection of friends who smoke and drink. Another interpretation for positive sibling effects can be seen through a "supply effect" wherein increased substance use by the older child provides greater opportunities for the younger sibling to smoke or drink through an implicit re-

[11] Maternal employment status could be used as a proxy for time constraints as it would reflect the mother's physical presence at home. However, lack of data on this variable precludes its measurement in my model.

[12] A recent paper by Price (2008) investigated the effects of birth order on parental time investments in children between the ages of 4 and 17 from two-children households, using data from the American Time Use Survey. He showed that at each point in time, parents make equal time investments in each sibling. Price's empirical findings lend support for my identification assumption. Moreover, he argued that birth order differences widen with large age gaps between children. Since siblings in my sample are on average spaced quite close together, economies of scale may encourage equality of parental investments across both children.

duction in the marginal cost or effective price of consuming the substance.[13] Given these many channels of influence, the true process underlying sibling effects may be a complex combination of these mechanisms and identifying them separately is beyond the scope of the current paper and my data.

I also estimate sibling effects parameterized by the age gap between siblings as well as the gender composition of the sibship. As expected, I find effects of larger magnitude for sibling pairs who are more closely spaced together. While sibling effects are significant and positive for same as well as mixed sex pairs, I find that they are larger among the latter. This is an interesting and somewhat surprising result that should be investigated with alternative datasets. Lastly, estimating sibling influences based on the gender of the younger sibling indicates stronger effects of the older sibling's behavior on younger girls as compared to younger boys.[14]

The results of this paper indicate the presence of opportunities for resource constrained parents to invest efficiently in favor of their older child, to reduce "bad" behavior. In the absence of sibling influences, such a policy biasing investments in favor of the firstborn can have detrimental effects on the behavior of the other child. However, positive sibling effects imply that curtailing the older child's behavior in this way can have spillover effects on the younger sibling, leading to greater payoffs to parents in overall child quality. These sibling effects are also of interest to public policies that aim to curtail adolescent substance use (for example, a legal drinking age law). The creation of within family "social multipliers" can amplify the effect of even a small policy intervention, by propagating its effect through siblings within a family.

The remainder of the paper proceeds as follows. Section 2 describes the various mechanisms underlying sibling influences. Sections 3 presents a basic conceptual framework to motivate the empirical model. Section 4 describes the data used in the empirical analysis. Section 5 outlines the econometric model and identification assumptions. Section 6 summarizes and discusses the empirical results. Finally, Section 7 concludes and presents avenues for further research.

[13] These mechanisms are discussed in greater detail in the next section.

[14] While it would be interesting to examine whether older siblings' influences vary by their gender, I am constrained from doing this owing to power issues. Estimating these effects conditional on whether the older child is a boy or girl would entail an exactly identified model, with only the age of the older sibling available as an identifying instrument. Since the standard errors from such a regression are quite large, no meaningful inferences can be made.

2 Mechanisms of Sibling Influences

Several studies in sociology and psychology have investigated the importance of the family and social environment in shaping a child's attitudes and behaviors. In particular, they emphasize siblings as being a significant domain of influence, especially in areas of risky behavior and substance use that accelerate during adolescence. The majority of these studies have found a positive association between siblings' smoking or drinking. If these effects are causal, then they imply that increased substance use by the older sibling leads to the younger sibling also engaging more in the same behavior. An older sibling's smoking or drinking can exacerbate a younger adolescent's tendency to engage in the corresponding behavior in more ways than one. In what follows, I discuss the various mechanisms that can underlie positive sibling effects.

Increased substance use by an older sibling can raise the marginal benefit of the same activity for the younger child. These preference shifts on the one hand can occur through the older sibling serving as some sort of mentor or "role model". Kuhn (1964) defines "orientational others" or "reference persons" as role models to whom subjects are committed emotionally and psychologically, and as a result try to model their own behaviors by directly replicating the behaviors of these significant others. Since siblings share the same environment, older brothers and sisters have opportunities to practice "drug advocacy", where they encourage substance use by their younger counterparts (Brook et al, 1990). Conversely, by reducing his cigarette or alcohol consumption, an older sibling may be a source of information concerning the unacceptability of these undesirable behaviors, and consequently exert corrective influences on the younger sibling leading him to reduce his consumption as well. This channel of interpretation is consistent with social learning models in which individuals are more influenced by people whom they believe are similar to themselves.

Another related mechanism through which the marginal benefit of smoking or drinking by the younger child can get affected is via peers - the selection of friends who engage in similar behaviors. Increased substance use by the older child may reflect his normative beliefs concerning the acceptability of smoking and drinking and serve to legitimize association with others who smoke or drink, outside the family. Then, if the younger sibling uses his older brother or sister as

references while shaping his own beliefs, he will find it acceptable or even desirable to associate with like-minded deviant non-family peers, thereby exacerbating his consumption of alcohol or cigarettes. Indirect effects through peer groups can also occur if an older sibling introduces his younger brother or sister, particularly those who are close in age, directly into his own friends circles consisting of individuals habituated to substance use.

Positive influences from the older to the younger child can also manifest through a "supply effect", wherein increased substance use by an older sibling can provide greater opportunities for the younger child to engage in the same behavior by altering the effective price the younger sibling would pay to smoke or drink. To economists, the price of the substance does not just include its monetary cost, but also the value of time spent obtaining cigarettes or alcohol, along with the expected penalities for underage drinking. On the one hand, direct contact between older and younger siblings makes it possible for older children to "hand over" cigarettes or bottles of beer to their younger brothers and sisters. On the other hand, adolescents can be exposed to their older brothers and sisters attitudes and behaviors even if they do not directly participate with their siblings, for example, through the theft of these substances from around the house or at parties of their older sibling. The more an older sibling smokes or drinks, the greater is the likelihood of detection and theft by the younger sibling. In either case, substance use by the older child results in a reduction in the marginal cost of consuming the same substance for the younger sibling, facilitating greater use of cigarettes or alcohol.

Overall, these mechanisms provide a theoretical basis for believing that smoking or drinking by an older sibling can exert causal effects on the corresponding behavior of his or her younger brother or sister. However, without appropriate data, it is very difficult to isolate these channels and measure their relative contribution. The current paper treats the sibling effect as a reduced form parameter, and I do not attempt to identify these specific mechanisms that underlie it.

3 The Conceptual Framework

An adolescent makes the choice to engage in substance use by comparing the perceived benefits with the perceived costs of smoking or drinking. I focus on the younger sibling (indexed by y) from

a two children family. In this context, the utility of the individual will depend on the consumption of the substance in question (say alcohol Y_y) as well as the consumption of all other goods C_y. Utility is assumed to be separable in these two components of consumption. Then, the younger sibling's problem is to maximize this utility subject to a budget constraint as follows:

$$\max_{C_y, Y_y} U(C_y) + e^{(\beta W_y + \nu_y)} V(Y_y)$$

subject to

$$Inc_y = C_y + PY_y$$

such that

$$U' > 0, U'' < 0, V' > 0 \text{ and } V'' < 0$$

In the above, the individual's income is denoted by Inc_y and the price of alcohol is *P*. I allow the marginal utility of alcohol consumption for the younger child to be affected by a vector of factors $W_y = f(X_y, Fam, I_y, Y_o)$. These factors include individual specific characteristics X_y (such as age and gender) and observable family background indicators *Fam* that are shared between siblings. In addition, I introduce two components that are believed to affect the marginal utility of drinking for the younger sibling. The first refers to investments made by parents in the younger child I_y. I also allow the younger child's marginal utility to depend on the current period drinking of other individuals in his peer group. In this case, I define the reference group of the younger sibling to be his older brother or sister whose alcohol consumption is denoted as Y_o. An individual's unobservable "taste" for drinking is captured by the term ν_y.[15] The parameter vector β assigns weights to the utility obtained from drinking for each individual according to his personal characteristics, family factors, investments and older sibling's behavior.[16] In this model, I abstract from all reciprocal influences from the younger to the older sibling.[17]

[15] It should be noted that ν_y can incorporate factors that are entirely individual specific (eg., depression) and/or unobservables that are shared between children in the same family. For example, the preference to drink can depend on whether the parents also engage in alcohol consumption, or whether the siblings go to a low quality school.

[16] See Pacula (1998).

[17] Support for this is stated in footnote 3 in the Introduction.

From the above model, we can derive the first order conditions for the younger child's maximization problem as,

$$[C_y]: \quad \frac{\partial U}{\partial C_y} - \lambda = 0 \tag{1}$$

$$[Y_y]: \quad e^{(\beta W_y + \nu_y)} \frac{\partial V}{\partial Y_y} - \lambda P \leq 0 \tag{2}$$

$$\text{Budget Constraint}: \quad Inc_y = C_y + PY_y \tag{3}$$

Note that Equation 2 is a Kuhn-Tucker condition that allows for a corner solution for Y_y. Hence, if the younger sibling does not engage in alcohol consumption, Equation 2 will hold with strict inequality and imply,

$$e^{(\beta W_y + \nu_y)} \left(\frac{MU_{Y_y}}{MU_{C_y}} \right) \Bigg|_{Y_y = 0} < P$$

The above inequality implies that if the marginal cost of engaging in alcohol consumption (captured by price *P*) is greater than the marginal benefit of such consumption, we will have $Y_y = 0$. For an interior solution ($Y_y \neq 0$), Equation 2 will hold with strict equality and we get the standard optimality condition where the younger sibling's choice of (C_y, Y_y) is determined at the point of equality between marginal cost and marginal benefit, i.e.,

$$e^{(\beta W_y + \nu_y)} \left(\frac{MU_{Y_y}}{MU_{C_y}} \right) = P$$

Taking logarithm of both sides, we get,

$$\beta W_y + \nu_y + \ln \left(\frac{MU_{Y_y}}{MU_{C_y}} \right) = \ln P$$

In principle, if *U* and *V* are continuously differentiable with a non-zero Jacobian, we can solve for the optimal consumption of the younger child as,

$$Y_y^* = f(\ln P, X_y, Fam, I_y, Y_o, \nu_y) \tag{4}$$

The above function can be linearized to obtain a reduced form regression of the younger child's behavior that depends on his individual characteristics, the common family environment, investments made in him and his older sibling's corresponding behavior. The question of interest in this paper is then to investigate the magnitude and sign of $\frac{\partial Y_y}{\partial Y_o}$, which I hypothesize is positive. This suggests that increased drinking (or smoking) by the older sibling leads to the younger sibling doing more of the same behavior, all else equal.

4 Data

The data used in this paper consists of a sample of sibling pairs from the National Longitudinal Survey of Youth 1997 Cohort (NLSY97). The initial wave of the NLSY97 consisted of a nationally representative sample of approximately 9,000 youths who were between the ages of 12 and 16 years as of December 31st, 1996. Respondents of this survey have been interviewed on an annual basis since the first round in 1997. Since I am interested in estimating sibling effects for young adolescents, the age distribution of the NLSY97 is well suited for my purpose.[18] An important feature of the NLSY surveys is that each child in the sibling pairs is independently surveyed. In addition to schooling and labor market characteristics, the NLSY97 contains detailed information

[18] Although siblings in this survey are young enough to be a part of my sample, one limitation of the NLSY97 is that the maximum age gap between siblings is five years. In contrast, the National Longitudinal Survey of Youth 1979 Cohort (NLSY79) or its Child and Young Adult (CYA) Supplement have a less restricted range of age differences. The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14 to 22 years old when they were first surveyed in 1979. The CYA Supplement is a biennial survey that began in 1986 and contains information on children born to the NLSY79 female respondents. A disadvantage with the NLSY79 survey is that alcohol and smoking related questions are asked of respondents only from 1982 or after, during which all respondents are already 18 years or older when many would have left home. With the CYA Supplement, most of the respondents are very young (less than 12 years of age) and so data on substance use behavior may not be very informative. Moreover, the questions are not consistently worded across all years.

on family background, personal demographics and a variety of behaviors that can be considered risky and delinquent.

The current study uses sibling pairs who are drawn from two children households.[19] I use data from 1997 through to 2006 and limit my sample to individuals between 12 and 18 years of age with non-missing information on the variables of interest. The outcome measures that I use for substance use are two continuous variables - the number of days the individual smoked cigarettes in the last 30 days and the number of days the individual drank alcohol in the last 30 days. In addition to sibling specific controls such as age, gender and mother's age at birth, I also include a vector of shared family factors representing race, region of residence, categories of mother's education, family income, household size, family structure and ethnicity.[20] My final sample for analyzing smoking and drinking behavior consists of 4227 sibling-year observations.

Table 1 presents unweighted sample means for the variables used in the empirical analysis. Panel A summarizes the data for the outcome measures of smoking and drinking. Panel B presents means for the explanatory variables used in the regressions. Across both the outcomes of substance use, the mean is higher among older siblings than younger siblings. The sample statistics for the control variables are the same across both the samples. The sample is split roughly equally between same and mixed gender sibling pairs. The mean age among older siblings is approximately 15 years while that among their younger counterparts is about 13 years. This makes siblings in my sample a little closer in age as compared to average American siblings (see Widmer (1997)). Across both the samples, about 66% live in urban areas and about 69% are white. The mean age of mothers when they gave birth to the younger child is about 25 years. About 34% of mothers have a high

[19] In the first wave of the NLSY97, roughly about 83% of multiple siblings belonged to two respondent households. It is possible that these households contained other siblings who were excluded from the first wave because they were either less than 12 or more than 16 years of age at the end of 1996. To ascertain that there were no other non-responding siblings in the family, I restricted my sample to those households which only contained two responding siblings and zero non-respondents.

[20] Since I only include two children households, household size does not refer to the sibship size, but to the number of people residing in the household other than the two children and their parents. I account for family structure with a dummy variable indicating an intact family consisting of both parents. Gender is represented by a dummy variable that is equal to 1 if male and 0 if female. In the NLSY97, family income is missing for many respondents in the sample. To avoid loss of information, I impute family income from available data for similar families. To impute family income, it is assumed that this variable is made up of two parts, one that can be explained by demographic characteristics of the family and the other that follows a stochastic process. The correlation between actual and imputed family income is roughly 90%.

school graduate degree. The mean household size is around 4 and about 60% of children live with both their parents. Mean family income in the sample is about $65,000. Lastly, a small portion across both samples, 10%, is of Hispanic origin.[21]

Table 2 presents these sample means conditional on substance use. I define indicator variables that are equal to 1 if the individual smoked cigarettes or drank alcohol in the last month, and 0 otherwise. The first two columns present these means for the subsamples of non-smokers and smokers, and the last two columns for non-drinkers and drinkers respectively. There are some clear differences in family characteristics when we compare individuals who engage in substance use with those who do not. Individuals who smoke are more likely to live in urban areas and be white than those who do not.[22] Individuals who smoke are more likely to have mothers at the extremes of the educational distribution - those who are high school dropouts, and mothers who are educated till the college level. However, individuals who do not smoke are more likely to have mothers with college plus level of education as compared to individuals who do smoke. Individuals who engage in substance use come from families with lower average income levels as compared to adolescents who do not smoke or drink. Studies have usually found that smoking and drinking by teenagers is more prevalent among those with low socioeconomic status parents. Teens who smoke are also more likely to come from a broken family as compared to their counterparts who do not. This is consistent with prior research that has compared adolescent smoking and drinking in two versus one-parent families (see Oman et al (2007), Kirby (2002)). For drinking, we find that most of the patterns remain as they did for smoking when the sample of drinkers is compared to non-

[21]The adolescent population (aged 10 to 19 years) is more racially/ethnically diverse than the overall population. While white and non-Hispanics comprise a majority of both populations, the adolescent population has a greater percentage of blacks and Hispanics than the population as a whole (US Census Bureau, 2003). According to the 2000 Census, about 63% of the adolescent population are white, and about 15% comprise of Hispanics. My sample somewhat overrepresents whites and underrepresents Hispanics. In terms of region of residence, about 80% of all adolescents aged 12 to 17 years live in suburban or central city settings, while the remaining 19% live in rural areas. This is higher than the percentage of individuals who live in urban areas in my sample. The Census reports that in 2002, about two-thirds of adolescents aged 12 to 17 years live with both parents, a figure that is somewhat higher than what I find in my sample.

[22]Most studies of adolescent substance use have found that black youths are less likely to smoke cigarettes than white youths. These studies have posited many reasons to explain this pattern. One is that black adolescents may be more likely to believe that tobacco products are being marketed specifically to them (McIntosh, 1995). Another is that black females may be less likely to use smoking as a weight-control strategy (Camp et al, 1993). Lastly, black youths may be less likely to consider cigarette smoking to be fun.

drinkers. Since these factors affect the probability that an adolescent engages in such behaviors, it is important that they be included in the regression equation of the younger child to obtain more accurate estimates of the sibling effect.

In Table 3, I tabulate sibling correlations in smoking and drinking by computing conditional probabilities of younger siblings' substance use, given their older siblings' choices. The top panel presents these probabilities for smoking, while the bottom panel presents the results for drinking. From the table, we can observe that younger siblings are consistently more likely to smoke or drink if they have older siblings who engage in the same behavior. The likelihood of smoking by the younger child when he has an older sibling who does not smoke is about 12%, and this probability increases to about 38% when the younger child has an older brother or sister who also smokes. To test if the probability of the younger child smoking is related to his older sibling's choice, I conduct the chi-square test of independence which strongly rejects the null hypothesis of independence. The patterns of these conditional probabilities are quite similar for the drinking outcome. These significant relationships suggest that older siblings may be affecting their younger sibling's behaviors. However, since siblings share many influences, these numbers should merely be interpreted as correlations and not as causal.

5 The Econometric Model

A regression equation for the younger sibling derived from Equation 4 can be written as,

$$Y_{yi} = \beta_0 + \beta_1 X_{yi} + \beta_2 Fam_i + \gamma I_{yi} + \theta Y_{oi} + \nu_{yi} \tag{5}$$

Equation 5 above posits that the smoking or drinking of the younger child, Y_{yi}, from a sibling pair *i* is a function of observable characteristics X_{yi} that vary between siblings *within* a household, Fam_i which is a vector of observable characteristics that vary *between* households (common family environment), parental investments I_{yi}, the older sibling's substance use Y_{oi} and an individual specific disturbance term ν_{yi}. The vector X_{yi} contains the gender and age of the younger child,

entered individually and as an interaction term. It also contains a variable measuring the mother's age at birth of the child. The common family factors included in the above regression are household size, region of residence, categories of mother's education, race, family income, ethnicity and family structure represented by two parent households.

The main coefficient of interest is the sibling effect parameter, θ, which captures the impact of the older sibling's behavior Y_{oi} on that of his younger counterpart, Y_{yi}. In this equation, there are two sources of endogeneity in the older sibling's behavior. First, there can be unobserved differences between families that affect the "taste" for smoking or drinking (e.g., parents in some families may smoke more than parents in others).[23] In addition, although parental investments I_{yi} are an important determinant of the younger sibling's smoking and drinking, they are difficult to control as they are multi-dimensional, encompassing monetary as well as non-monetary measures. If investments in the younger child were independent of the older sibling's behavior, omitting I_{yi} would not induce any bias in the estimate of the coefficient θ. However, parental decisions to invest in their offspring may be endogeneous to child behavior, leading us to suspect that the correlation between omitted investments and Y_{oi} is not necessarily zero. We estimate,

$$Y_{yi} = \beta_0 + \beta_1 X_{yi} + \beta_2 Fam_i + \theta Y_{oi} + \epsilon_{yi} \tag{6}$$

where

$$\epsilon_{yi} = \gamma I_{yi} + \nu_{yi}$$

Since $Cov(Y_{oi}, \epsilon_{yi}) \neq 0$, standard regression techniques such as Ordinary Least Squares (OLS) that fail to control for this endogeneity will produce biased estimates of the sibling effect parameter θ. To address endogeneity, I make use of instrumental variables (IV) estimation. The next section describes my instruments and identification assumptions.

[23] Price is also uncontrolled for in this regression.

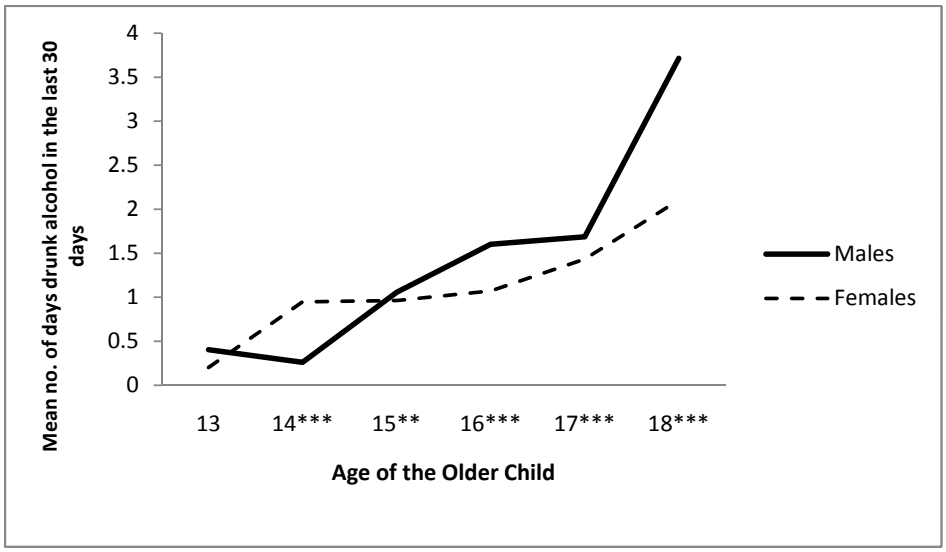
5.1 Instrumental Variable Estimation

To account for endogeneity in the older sibling's behavior, my instruments use the demographic characteristics of the *older* child as a source of variation in Y_{oi}. Specifically, my instruments use gender differences in smoking and drinking across age among older siblings. Figure 2 plots the consumption of alcohol and cigarettes among older boys and girls by age, controlling for observable characteristics. I estimate a simple regression of the older child's outcome on his own characteristics (age and gender), as well as the family specific variables used in Equation 5 of the younger child. The top and bottom panels of Figure 2 plot regression adjusted estimates of the number of days on which older siblings aged 13 to 18 years from two children families drank alcohol and smoked cigarettes respectively, in the last month prior to the survey, by gender. The dotted line represents the mean female outcome while the solid line represents the mean male outcome. This figure also shows gender differences in smoking and drinking across age which are statistically significant at almost every age of the older child. In the early teenage years, boys and girls have similar average rates of alcohol consumption, but males drink consistently more than females in mid to late adolescence. The patterns are not quite as clear for the smoking outcome. At younger ages, gender differences are larger for smoking than drinking. In the later teenage years, although the gender gap in smoking is not as striking as what it was for drinking, it still appears that the frequency of smoking is somewhat higher for boys as compared to girls. This pattern is confirmed in the first stage estimates for the smoking outcome.[24]

The observation that gender differences exist among adolescents in cigarette and alcohol use is supported by many studies in the substance use literature. For example, Lewinsohn et al (1996) used a survey dataset of 1709 participants randomly selected from nine senior high schools in West Oregon and found that males not only had greater frequency of alcohol consumption, but also consumed a greater quantity of alcohol per drinking episode. Barnes et al (2002) used a dataset of 19,321 New York State seventh to twelfth grade students and found that females have lower levels of daily alcohol consumption as compared to males. These observations are consistent

[24]The coefficient on the interaction between the older child's age and gender in the first stage regression for smoking is positive and significant, suggesting that as older siblings age, boys overtake girls in cigarette consumption (see Table 10).

Figure 2 : Gender Differences in Smoking and Drinking Among Older Children in the NLSY97 : Conditional Data



*, ** and *** denote significance of age-wise gender differences at 1%, 5% and 10% levels respectively.

with other findings such as Johnston, O'Malley and Bachman (2000) who found that males show higher levels of alcohol consumption, binge drinking and overall illicit drug use than females. With respect to smoking too, Vidrine et al (2006) found that adolescent boys were more likely to be current smokers as compared to adolescent girls. The notion that gender differences are not constant at all ages are supported in the Minnesota Student Survey report, which depicts that the extent to which boys and girls differ in their past month smoking and drinking varies across sixth, ninth and twelfth grade students. The findings of this survey support the observation from Figure 1 that gender differences in drinking mostly emerge at higher ages among adolescents, and boys and girls have similar rates of alcohol consumption during early adolescence. Although researchers have also found that age is positively related to substance use among adolescents (Ardelt and Day (2002)) the current study is the first to make use of gender variation across age in cigarette and alcohol use as instruments to predict the older sibling's endogenous outcome.

Peer pressure is usually given heightened importance in discussions of adolescent misbehavior and is implicated in many instances of risky behavior in which adolescents engage such as substance use and delinquency. Most studies have indicated that susceptibility to peer pressure usually follows an inverted U shaped pattern, peaking at about age 14 and declining thereafter.[25] However, this pattern might not be the same for boys and girls. For male adolescents, membership in a deviant group that promotes qualities considered traditionally masculine such as toughness may be a way to establish a masculine identity and increase their status, reputation and physical attractiveness as "men" prior to entering the adult world (Hunt, Joe and Waldorf, 1996., Matt, 1999). Since boys usually mature later than girls, they may become conscious of their physicality more during mid to late adolescence and have lower resistance to peer pressure as compared to girls in those age groups. This perspective might explain why gender differences in substance use are not constant at all ages.

With these instruments, we can presumably control for confounding effects arising out of most unobservables such as substance price and family background characteristics such as parental smoking or drinking. There is no apparent reason to believe that the older child's characteristics should be correlated with these variables. With respect to parental investments, the instruments

[25] See Berndt (1979), Krosnick and Judd (1982), Steinberg and Silverberg (1986).

are valid as long as investments made in the younger child are independent of the older child's age and gender. In other words, I allow a younger sibling's own characteristics to affect parental investments in him, and can control for omitted investments if the older sibling's characteristics do not affect investments made in the younger child. It is possible that this assumption may not be necessarily true if parents face time or income constraints. In my regressions, I control for the main effects of family income. Across-family differences with respect to constraints will therefore only be generated by differences between groups in the "costs" of spending time with children. It should be noted that I control for a wide range of observables that place families in different structural positions such as family structure and socioeconomic status. Assuming that these variables are a reasonable approximation of parental labor supply, even if there are remaining unobservable between-family differences in work or time constraints, I argue that they may be small. In any case, rather than just invoking the assumption, I use data on parental investments available in the NLSY97 to test the identification assumption, providing empirical support for its validity.

My strategy allows me to estimate sibling effects that depend on the gender composition of the sibship and the age gap between the siblings. To estimate sibling effects separately based on the gender mix of the siblings, I interact the older child's outcome in Equation 6 with the dummy variables $SameSex_i$ and $MixedSex_i$, representing same sex and mixed sex sibling pairs respectively. Then,

$$Y_{yi} = \beta_0 + \beta_1 X_{yi} + \beta_2 Fam_i + \theta_1(Y_{oi} * SameSex_i) + \theta_2(Y_{oi} * MixedSex_i) + \epsilon_{yi}$$

To estimate the sibling effect conditional on the age gap between siblings, I define an indicator variable $BigAgeGap_i$ that takes value 1 if the age gap between a pair of siblings is greater than the average age difference between them. From the equation below, θ is the sibling effect for those pairs with a below average age gap between them, and $\tilde{\theta}$ represents the *difference* in sibling effects between pairs with an above average and below average age difference. If we believe that sibling effects are stronger in pairs that are closer in age to one another, then the estimate of $\tilde{\theta}$ should be negative in sign.

$$Y_{yi} = \beta_0 + \beta_1 X_{yi} + \beta_2 Fam_i + \theta Y_{oi} + \tilde{\theta}(Y_{oi} * BigAgeGap_i) + \epsilon_{yi}$$

5.2 Testing the Validity of the Identification Assumption

Parental investments are an important determinant of adolescent behavior, yet it is difficult to quantify them in entirety with existing datasets. This is because they can be multi-dimensional, encompassing monetary as well as non monetary measures. Investments can be *economic* in nature through parental spending on child needs, or in the form of transfers or allowances given to the child. Investments can also be *interactional* or *supervisory* through parental involvement in their child's life. Parents can be involved in their child's academic life by regularly talking to them about issues related to school activities and class performance. In addition, parental knowledge about what the child does out of school, who the child's friends are and how the child spends his or her money may also have a bearing on their behavior. Lastly, some researchers have emphasized the role played by *cultural* investments wherein parents provide their children with the opportunity to participate in cultural activities such as going to museums and enrolling in art, music and dance. Such types of investment in the child's cultural capital are known to positively affect the mental and physical well being of young adolescents (Bourdieu, 1977).

Models of intrafamily allocation suggest that parental preferences for investing in their offspring can be driven by child characteristics. On the one hand, parental responses can result out of a pure gender or birth order preference which is a feature of the parental utility function. Boys and girls can also have distinct production functions, which can lead to either real or perceived differences in the returns to parental inputs or the "price" of raising a boy versus a girl (eg., boys demand more time in general or boys need fathers more than girls).[26] While the younger sibling's

[26]Although discriminatory investments in offspring motivated by child gender is a theoretical possibility, there is not much empirical evidence for an overt gender bias in resources allocated to boys and girls in industrialized societies such as the US (Lundberg, 2005). In wealthy societies, parental dependence on support from their progeny in old age has eroded owing to a well established pension system. Thus, motives for a pronounced male preference have diminished to a considerable extent. Falling fertility and rising female labor force participation have been responsible for more emancipation in the perception of women. Since the differences in economic costs and material returns to sons and daughters appear minimal, incentives to invest differently in offspring by gender are less of an issue in recent

own characteristics can affect his or her investments, the crucial assumption I need for consistency of the IV estimation is that there is no direct relationship between the *older* sibling's age and gender and parental investments made in the younger child. I formalize this identification assumption through a standard bias derivation, and then proceed to describe the NLSY97 investment data that I use to test it.

In order for the IV estimates to be consistent, the correlation between $\hat{Y_o}$ and I_y should be zero, controlling for all else. In Appendix 1, I use standard econometrics to formalize the manner in which we can investigate empirically the validity of this assumption. The bias derivation shows that in order to test the identification assumption, we have to estimate a reduced form equation of the younger sibling's investments regressed on all the instruments (excluded and included), and test the joint significance of the excluded instruments. I do this with data from the NLSY97 on variables that approximate parental investments.[27]

The NLSY97 asks a few questions on parent child associations that reflect the degree of involvement of a parent in their child's school and personal life. These are asked separately of the mother and the father. There are also questions in this survey pertaining to monetary transfers captured by allowances given by the parent to the child. In particular, I make use of the following thirteen investment measures from the survey[28]:

1. The individual's mother (father) helps him or her with what is important to the individual
2. The individual's mother (father) knows who he or she is with when not at home
3. The individual's mother (father) knows his or her teachers and school activities
4. The individual's mother (father) knows his or her close friends
5. Degree of parental monitoring of the individual by his or her mother (father)

times.

[27] I choose to go this route of utilizing the data to only test my assumption rather than directly in my regression. By quantifying investments with this data in the main equation of interest, the problem of poor measurability persists. Also, since investments are presumably endogenous to child behavior, we would have to find suitable instruments to proxy them.

[28] The first 6 questions are asked separately of mothers and fathers, amounting to 12 investment measures. The allowance question is asked of the family in general.

6. Degree of how supportive is the relationship of the individual with his or her mother (father)

7. Total allowance the individual received from the family in the past year

Questions 1 through 4 are Likert Scale variables in the survey, and appear as five-point scales with options ranging from "Strongly Disagree" to "Strongly Agree". Questions 5 through 7 are continuous variables. In order to convert the Likert Scale data to the nominal level, I combine all agree and disagree responses into dummy variables representing two categories of "Often" and "Rarely". For example, for the first question, I assign a value 1 to the responses of "Agree" and "Strongly agree" categorizing them as "Helping Often", and assign 0 to the remaining responses to categorize them as "Helping Rarely". I use the same conversion for questions 2 through 4. For Questions 5 and 6 representing the degree of monitoring and a supportive relationship, I create dummy variables that represent "Above Average Monitoring" and an "Above Average Supportive Relationship".[29]

Table 13 in Appendix 2 reports the results for the test of joint significance of the excluded instruments in the auxilliary regression of the younger sibling's investments. A failure to reject the null hypothesis indicates that there is no significant relationship between the younger child's investments and the older sibling's characteristics that are used as instruments. From the table, it can be observed that in all the thirteen cases, I fail to reject the null hypothesis at the 1% and 5% levels of significance, for both mothers and fathers. In twelve of the cases, I fail to reject the null hypothesis at the 10% level of statistical significance. Since these results are based on a particular conversion of the Likert Scale variable which may be ad-hoc, I test to see if the identification assumption is robust under a set of alternative definitions for the investment dummies and obtain largely similar results. I do not report these alternative results in the paper but they are available upon request. Although these data do not wholly capture all dimensions of parental investments, to the extent that other unobserved components display similar distributional properties, it is reasonable to expect that my IVs are not correlated with investments made in the younger sibling.[30]

[29] I define these variables in this way since it is reasonable to expect that at some basic level, mothers and fathers will monitor and support their children equally.

[30] I also test this assumption for subsamples based upon gender composition. First, I investigate whether the identi-

6 Estimation Results

Table 4 reports the results from naive models that use OLS to estimate the effect of the older sibling's substance use on his younger sibling's corresponding behavior. The first and third column report the magnitude of the sibling coefficient from a regression that controls only for the younger child's demographic characteristics, for smoking and drinking respectively. Columns 2 and 4 report the results from regressions that include family characteristics as well. The table reports that smoking and drinking by an older sibling are strongly positively associated with a younger sibling's behavior, even after controlling for family characteristics. The positive coefficient on age implies that younger siblings smoke and drink more as they age. The significant and positive sign of the coefficient of the interaction term indicates that as younger siblings age, males engage more in both types of substance use as compared to females. With respect to family factors, white individuals smoke and drink significantly more than those from black families. Individuals with older mother's engage more in substance use, but the coefficient on mother's age is not statistically significant. Living in an urban area increases the incidence of both types of substance use. Mother's education is mostly positively associated with substance use, but is not very significant for drinking. The difference in the substance use of adolescents from large and small families is not statistically significant. Family structure seems to be an important predictor of adolescent substance use, with lower smoking and drinking among children from intact families. Lastly, being Hispanic is negatively related to substance use by teens, and is significant in most cases.

Although the older sibling's use of alcohol and cigarettes is strongly positively correlated with his younger counterpart's corresponding behavior, these sibling "effects" do not capture any causal influences. As indicated, siblings can share many common factors that are unobserved and uncontrolled for in these regressions. Tables 11 and 12 in Appendix 2 report the first stage results from 2SLS regressions for smoking and drinking respectively, across a range of specifications. I use the age and gender of the older sibling, entered individually and as an interaction term, as excluded

fication assumption holds within same and mixed gender sibling pairs. Second, I examine whether the instruments are uncorrelated with younger child investments in families with younger girls versus younger boys. Even within these subgroups, I am able to find consistent evidence in favor of no significant correlation between the younger child's investments and the older sibling's characteristics. I do not report these results for brevity but they are available upon request.

instruments. In these tables, Column 1 reports the first stage estimates for a regression of the older sibling's smoking (drinking) on the corresponding behavior of the younger child for the pooled sample. Columns 2 and 3 do the same based on the gender mix of the sibship for same and mixed sex pairs respectively. Column 4 reports the first stage results for cross substance regressions, wherein I regress the younger sibling's smoking (drinking) on the older sibling's smoking *and* drinking.[31] For each of the two outcomes, the results are qualitatively similar across all specifications. It can be observed that not only are the instruments individually statistically significant, but the Staiger and Stock First Stage F-Statistic for weak instruments is also well above the accepted level of 10 in all regressions. These diagnostics suggest that the estimated specifications do not appear to suffer from problems associated with weak instruments. The effects of the other control variables are for the most part, similar to what was obtained in Table 4.

Table 5 reports the second stage estimates for the pooled sample consisting of the entire constellation of all sibling pairs. Column 1 states the results for smoking and Column 2 for drinking. The estimates for the sibling effect are reported at the top of the table, and are highly significant and positive in sign. To get a sense of the magnitude of the coefficients, I report the effect of the older sibling smoking or drinking on the corresponding behavior of the younger sibling as a proportionate change at the bottom of the table. The effect of the older sibling's smoking behavior on the younger child is about 72%. The same effect for drinking is roughly 30%. Thus, the influence of the older child's behavior on his younger sibling is much stronger for the case of smoking than it is for drinking.[32]The coefficients of the other demographic and family background variables are mostly qualitatively similar to those obtained for the older child in the first stage regressions.

In Table 6 below, I report the magnitudes of the sibling coefficient by allowing the variables used as excluded instruments to differ across alternative specifications. The first row in the table reports the estimates of the sibling effect when the full set of instruments were used to predict the

[31]I only report the first stage estimates for the cross substance, as the results for the own substance are very similar to what is reported in Column 1 of the table.

[32]The fact that sibling effects are much stronger for smoking as compared to drinking provides some support in favor of the supply effect. It is presumably easier for both siblings to smoke cigarettes together as compared to consuming bottles of alcohol. Given that bottles of alcohol are harder to hide from parents and other caregivers as compared to cigarettes that are less conspicuous, this allows an older sibling to possess cigarettes and give them to his younger sibling more often as compared to bottles of beer (or for the younger sibling to steal cigarettes). Morover, the symptoms of drinking are easier to detect as compared to smoking.

older child's outcome. This regression assumed that neither the older sibling's age nor gender had any direct effect on a younger child's smoking and drinking. I relax this assumption to some extent in the next three rows, which report the estimates of the sibling coefficient for cases that allow the interaction variable of age and gender, the older sibling's age and the older sibling's gender respectively, to affect the younger child's behavior directly.[33] It can be seen that the magnitude of the sibling effect remains robust to these alternative specifications.[34]

In addition to regressions for the entire sample of all sibling types, I also estimate sibling effects conditional on the age gap and gender mix of siblings. The results of sibling effects parameterized by the age gap are reported in Table 7. To estimate sibling influences conditional on the age difference, I interact the older sibling's smoking and drinking with a dummy variable indicating if the siblings belongs to a sibship with an above average age difference. The negative sign of these interaction terms are consistent with the interpretation that stronger sibling influences can be observed for pairs that are closer together in age. These effects are statistically significant and stronger among pairs that smoke.[35] I also estimate sibling effects by the gender composition of the sibship, for same and mixed sex sibling pairs, and report the 2SLS estimates in Table 8. These gender-mix influences are estimated by interacting the older child's outcome in the second stage regression with indicators representing same or opposite sex siblings. The sibling effects are statistically significant and positive in sign. I report the elasticities of the younger child's behavior in response to his older sibling's behavior at the bottom of the table. Specifically, if a same sex older sibling smokes more by 1%, his younger brother or sister responds by an increase of about

[33] As said earlier, one reason why the older sibling's gender or age can directly affect a younger sibling's smoking or drinking could be because of parental allocation decisions. On the one hand, investments in the younger child can change with the older sibling's characteristics because of gender or birth order preferences. They could also be affected through parental employment decisions, which are correlated with time spent by parents with their children. For example, Lundberg and Rose (2002) have shown that first born sons increase the labor supply of American fathers. Thus, these tests provide additional support for the identification assumption that the older sibling's characteristics has no direct effect on the younger sibling's behavior through omitted parental investments.

[34] Although not reported in the paper, I estimated 2SLS specifications that made even weaker identifying assumptions. For example, in one case I only used the gender of the older child as an excluded instrument, and allowed both the age and the interaction term to enter the second stage. The point estimate of the sibling coefficient in this exactly identified model was about 0.5 for smoking. However, it should be borne in mind that the exactly identified models are very demanding in terms of power, so in most cases the standard errors were too large for any meaningful inferences to be drawn.

[35] For brevity of space, I do not report the first stage results of the age gap effects, but they are available upon request.

Table 6 : Robustness of Instruments

Excluded Instruments	**Smoking**	**Drinking**
All Instruments (Male Old, Age Old, Male Old * Age Old)	**0.42*** (0.086)**	**0.17** (0.091)**
Age Old, Male Old	0.42*** (0.090)	0.20** (0.114)
Male Old, Male Old * Age Old	0.45*** (0.105)	0.18** (0.098)
Age Old, Male Old * Age Old	0.41*** (0.094)	0.19*** (0.075)
No. of Observations	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, younger child characteristics, time dummies and family factors.

0.6%. Similarly, if a mixed sex older sibling's drinking frequency increases by 1%, the younger child's corresponding consumption of alcohol goes up by roughly 0.4%. Although the sibling effects are positive for both same and mixed gender pairs, they are stronger for mixed sex pairs as compared to pairs of the same sex, an interesting yet surprising result.[36]

Additional regressions I estimate involve cross substance sibling effects. In other words, I regress the younger child's smoking (drinking) on his older sibling's smoking *and* drinking, in Table 9. It is interesting to examine these cross substance effects for two reasons. First, they may provide evidence either against or in favor of the supply effect. Second, depending on the relative magnitudes of the cross substance and own substance effects, it may present a stronger case for causality. While the effect of the older child smoking on the younger sibling's smoking is positive and significant, the cross substance effect is statistically insignificant. In the case of drinking, although both the older sibling's smoking and drinking exert significant effects on the younger sibling's drinking, the cross substance influence is very small as compared to the own substance effect.[37] As a final exercise, I allow sibling effects to vary by the gender of the younger child. The second stage estimates are reported separately for younger boys and girls in Table 10. Specifically, I find that although the sibling effects are positive and significant for younger boys and girls, the effect of the older sibling's behavior is considerably higher for younger girls as compared to boys, for both smoking as well as drinking.[38]

Given the many channels of sibling influence as described earlier on in the paper, the true process underlying sibling effects may be a complex combination of these many mechanisms, and

[36]Most studies in psychology and sociology that have estimated sibling effects by the gender composition of the sibship have either found no differences between same and mixed sex pairs, or larger effects for same sex pairs. While one cannot deny the fact that sibling effects are indeed significant and important in substance use behavior, the result that they are higher for mixed sex pairs should be at best treated with caution. More investigation is required to understand if this finding is true or an idiosyncratic result of the dataset used. I also parameterize sibling influences by the age gap within the sibship gender mix, and again find stronger sibling effects for pairs with a smaller birth spacing. These findings are not reported but are available upon request.

[37]I estimate these cross substance effects conditional on the age gap between siblings, and the findings again are in favor of stronger effects for pairs with smaller age differences between them. I do not report results in the paper, but they are available upon request.

[38]This finding is consistent with the generally held perception among sociologists and psychologists that adolescent girls usually report higher levels of intimacy with their peers (Blyth and Foster-Clark, 1987), derive their self-esteem from the reflected appraisals of others, and are more self conscious and other-oriented than boys (Rosenberg and Simmons, 1975).

identifying each of them separately is beyond the scope of the current paper. Stronger sibling effects between pairs close in age provide support for both the socialization and peer selection hypotheses. Children with a smaller birth spacing between them presumably live together in the same environment for longer as compared to siblings with large age differences. Such types of siblings are also likely to be more similar in their preferences and attitudes as compared to siblings spaced wide apart. This will in turn facilitate greater interactions between pairs closer in age. Closely spaced siblings are also more likely to share friends who promote substance use behaviors such as smoking and drinking. The finding that cross substance effects are either statistically insignificant (for smoking) or very small (for drinking) provides support in favor of the supply effect. The finding that a younger sibling smokes (drinks) only if has an older sibling who also smokes (drinks) suggests that he is probably obtaining cigarettes directly or stealing them from his older brother or sister. The insignificant cross substance effects also present a stronger case for causality, as it indicates that it is necessary for an older sibling to consume the same substance as the younger to exert a strong effect.

The finding that sibling effects are stronger for mixed gender pairs as compared to same gender pairs makes it less clear about the relative magnitudes of the various mechanisms. An interesting extension of the current study would be to integrate variables that measure the quality of the sibling relationship, in terms of overall friendliness or hostility. The association between the substance use of older and younger siblings might be stronger for adolescents who have a better relationship with one another, and who feel close to their older sibling. In contrast, individuals who do not get along with their older brothers or sisters may not be all that affected by their behavior. There have been a few suggestions in the psychology literature that same gender siblings tend to be more competitive with one another, while mixed sex pairs exhibit more camaraderie in their relationship. Although premature to conclude fully, inclusion of such variables may help throw some light on the secondary result in this paper that sibling effects for mixed sex pairs are larger than that for same sex pairs.

7 Conclusion

Although the role of the family in adolescent substance use has been a question of growing interest in psychology and sociology for several decades, an evaluation of the impact of siblings on behaviors such as smoking and drinking has been largely ignored by economists. This is especially surprising given the plethora of studies dealing with the estimation of effects exerted by non-family, school going peers. The current study attempts to fill this gap in the literature by studying sibling effects in adolescent substance use with data from the National Longitudinal Study of Youth 1997 (NLSY97). I define sibling effects to be the causal effect of the older sibling's smoking and drinking on the corresponding behavior of his younger sibling.

Estimating causal sibling effects necessitates accounting for endogeneity in the older sibling's behavior. This can result from across-family variability in family background that can affect the "taste" for consuming cigarettes or alcohol (for example, parental substance use), and from omitted parental investments leading to a correlation between siblings' behaviors. To control for endogeneity in the older child's outcome, I use differences in smoking and drinking by gender and age among older siblings as instrumental variables. The results point toward significant and positive sibling effects for smoking as well as drinking. Parameterizing sibling effects by the age gap between siblings, I find stronger effects between pairs that are more closely spaced to one another. In terms of the gender mix of the sibship, the effects are positive and significant for both same and mixed gender pairs, but stronger for mixed sex pairs. I also find that younger girls are on average influenced more by their older siblings as compared to younger boys.

These positive sibling effects are potentially relevant and interesting to studies that deal with household allocation of resources in multiple child families. Under the reasonable premise that parental investments improve child behavior, heterogenous investments (concentrated on the first-born) in the absence of sibling effects ($\theta = 0$) might have adverse consequences for the outcome of the younger child. In contrast, the presence of significant and positive sibling effects ($\theta > 0$) can provide parents, bound by time and financial constraints, an opportunity to invest in both children efficiently. In this case, it may be optimal for parents to increase the amount of investment in their first child to reduce his or her smoking and drinking. Positive sibling effects imply that

curtailing the older child's behavior in this way can have spillover effects on the younger sibling, leading to greater payoffs to parents in overall child quality. Naturally, the exact division of these investments will depend upon the relative magnitudes of the sibling effect and the effect of parental investments on an adolescent's behavior, an important arena for additional work. Nonetheless, the sheer existence of significant and positive sibling effects have interesting implications for parental investment strategies.

A secondary implication of positive sibling effects is that they may serve to amplify the effects of public policies. Since an adolescent's smoking and drinking is affected by the same behavior of his older sibling, any prevention program aimed at curtailing teenage substance use has both direct and indirect effects. For example, consider the introduction of a minimum legal drinking or smoking age. The immediate impact of this is that it reduces an adolescent's substance consumption. However, positive sibling effects mean that this intervention creates a social multiplier, wherein a lower consumption of alcohol and cigarettes by the first child will decrease the odds of smoking and drinking for younger adolescents within a family. When the social multiplier is large (for example, in closely spaced sibships), even small policy interventions may have large effects on individual behaviour.

While the current study has shed some light on the importance of siblings in determining individual behavior, it has also brought to fore the challenges involved in estimating causal effects, thereby outlining a substantial research agenda. More comprehensive data on parental investments will allow researchers to compare sibling influences to effects exerted by investments. The inclusion of variables that measure the quality and cohesiveness of the sibling relationship may help to understand why siblings from one type of sibship are affected more than others. Lastly, the sibling bond can become more significant over the life course on account of changing family structure caused by divorce, remarriage etc. Accounting for detailed family structure and various forms of parental insufficiency in sibling models is also a fruitful area for future research.

References

[1] Ardelt, M. and Day, L. 2002. "Parents, Siblings, and Peers: Close Social Relationships and Adolescent Deviance" *Journal of Early Adolescence* 22:310-49.

[2] Argys, Laura M; Rees, Daniel I; Averett, Susan L. and Witoonchart, Benjama. 2006. "Birth Order and Risky Adolescent Behavior" *Economic Inquiry* 44(2):215-23.

[3] Barnes, G.M; Welte, J.W; and Hoffman, J.H. 2002. "Relationship of Alcohol Use to Delinquency and Illicit Drug Use in Adolescents: Gender, Age and Racial/Ethnic Differences " *Journal of Drug Issues*:153-78.

[4] Behrman, Jere, Pollak, A. Robert and Paul Taubman. 1982. "Parental Preferences and Provision of Progeny" *Journal of Political Economy* 90(1):52-73.

[5] Berndt, T. 1979. "Developmental Changes in Conformity to Peers and Parents" *Developmental Psychology* 15: 608-16.

[6] Blyth, D. A., and F.S. Foster-Clark. 1987. "Gender Differences in Perceived Intimacy with Different Members of Adolescents Social Networks" *Sex Roles* 17: 689-719.

[7] Bourdieu, Jere R; Pollak, Robert and Taubman, Paul. 1990. "Parental Preferences and the Provision for Progeny" *Journal of Political Economy* 90(1):52-73.

[8] Boyle, M.H; Sanford, M; Szatmari, P; Merikangas, K. and Offord, D.R. 2001."Familial Influences on Substance Use by Adolescents and Young Adults" *Canadian Journal of Public*

Health 92(3)1:206-09.

[9] Brook, J.S; Whiteman, M; Gordon, A.S and Brook, D.W. 1990."The Role of Older Brothers in Younger Brothers' Drug Use Viewed in the Context of Parent and Peer Influences" *The Journal of Genetic Psychology* 151:59-75.

[10] Buhrmester, D. 1992. The Developmental Courses of Sibling and Peer Relationships. In F. Boer and J. Dunn (Eds.), *Children's Sibling Relationships: Developmental and Clinical Issues*. Hillsdale, NJ:Erlbaum:19-40.

[11] Camp, D.E; Klesges, R.C and Relyea, G. 1993. "The Relationship between Body Weight Concerns and Adolescent Smoking" *Health Psychology* 12:24-32.

[12] Clark, Andrew E and Loheac, Youenn. 2006. "It wasn't Me, it was Them! Social Influence in Risky Behavior by Adolescents" *Journal of Health Economics* 26(4):763-84.

[13] Cohen, D.A and Rice, J. 1997. "Parenting Styles, Adolescent Substance Use and Academic Achievement" *Journal of Drug Education* 27: 199-211.

[14] Dahl, G.B and E. Moretti. 2004. "The Demand for Sons: Evidence from Divorce, Fertility, and Shotgun Marriage" *NBER Working Paper* 10281, Cambridge, MA.

[15] Duncan, G; Boisjoly, J and Harris, K. 2001. "Sibling, Peer, Neighbor and Schoolmate Correlations as Indicators of the Importance of Context for Adolescent Development" *Demography* 38:437-47.

[16] Evans, W., W. Oats and R. Schwab. 1992. "Measuring Peer Group Effects: A Study of Teenage Behavior" *Journal of Political Economy* 100(5):966-91.

[17] Fagan, Abigail A. and Najman, Jake M. 2005. "The Relative Contributions of Parental and Sibling Substance Use to Adolescent Tobacco, Alcohol and Other Drug Use" *Journal of Drug Issues* 35(4):869-83.

[18] Gaviria, Alejandro and Raphael, Steven. 2001. "School Based Peer Effects and Juvenile Behavior" *The Review of Economics and Statistcs* 83(2):257-68.

[19] Haurin, R.J and Mott, F.L. 1990 "Adolescent Sexual Activity in the Family Context: The Impact of Older Siblings" *Demography* 27(4):537-57.

[20] Hunt, G., Joe, K., and Waldorf, D. 1996. "Drinking, Kicking Back and Gang Banging: Alcohol, Violence and Street Gangs" *Free Inquiry in Creative Sociology* 24:123-32.

[21] Jenkins Tucker, C., Barber, B. L., and Eccles, J. S. 1997. "Advice About Life Plans and Personal Problems in Late Adolescent Sibling Relationships" *Journal of Youth and Adolescence* 26:63-76.

[22] Johnston, L.D., O'Malley, P.M. and Bachman, J.G. 2000. National Survey Results on Drug Use from the Monitoring the Future study, 1975-1999, Vol. 2 *NIH Publication* No. 00-4803, Bethesda, MD: Department of Health and Human Services.

[23] Kirby, J. B. 2002. "The Influence of Parental Separation on Smoking Initiation in Adolescents" *Journal of Health and Social Behavior* 43:56-71.

[24] Krosnick, J. and Judd, C. 1982. "Transitions in Social Influence at Adolescence: Who Induces Cigarette Smoking?" *Developmental Psychology* 18:359-68.

[25] Kuhn, M.H. 1964. "The Reference Group Reconsidered" *The Sociological Quarterly* 5:5-23.

[26] Lamborn, S., Mounts, N., Steinberg, L. and Dornbusch, S. 1991. "Patterns of Competence and Adjustment among Adolescents from Authoritative, Authoritarian, Indulgent and Neglectful Homes" *Child Development* 62: 1049-65.

[27] Lewinsohn, P. M., Rohde, P., and Seeley, J. R. 1996. "Alcohol consumption in High School Adolescents: Frequency of Use and Dimensional Structure of Associated Problems" *Addiction* 91:375-90.

[28] Lundberg, Shelly. 2005. "Sons, Daughters and Parental Behavior" *Oxford Review of Economic Policy* 21(3):340-56.

[29] Lundberg, Shelly. and Rose, E. 2002. "The Effects of Sons and Daughters on Men's Labor Supply and Wages" *Review of Economics and Statistics* 84(2):251-68.

[30] Manski, C.F. 1993. "Identification of Endogeneous Social Effects: The Reflection Problem" *Review of Economic Studies* 60(3):531-42.

[31] Manski, C.F. 2000. "Economic Analysis of Social Interactions" *Journal of Economic Perspectives* 14(3):115-36.

[32] Matt, E. 1999. "Youth, Masculinities and Delinquency: Young Men Between Rituals of Masculinity and the Search for Autonomy" *Zeitschrift für Soziologie der Erziehung und Sozialisation* 19: 259-76.

[33] McIntosh, H. 1995. "Black Teens Not Smoking in Great Numbers" *Journal of The National Cancer Institute* 87(8): 564.

[34] Minnesota Departments of Education, Health, Human Services, and Public Safety. 2007. "Minnesota Student Survey: 1992-2007 Trends".

[35] National Institute on Drug Abuse. 2008. "Monitoring the Future: National Survey Results on Drug Use, 1975-2008", Volume I.

[36] Oettinger, Gerald S. 2000. "Sibling Similarity in High School Graduation Outcomes: Causal Interdependency or Unobserved Heterogeneity" *Southern Economic Journal* 66(3):631-48.

[37] Oman, R. F., Vesely, S. K., Tolma, E., Aspy, C. B., Rodine, S., and Marhsall, L. 2007. "Does Family Structure Matter in the Relationships Between Youth Assets and Youth Alcohol, Drug, and Tobacco Use?" *Journal of Research on Adolescence* 17:743-66.

[38] Pacula, R.L. 1998. "Does Increasing the Beer Tax Reduce Marijuana Consumption?" *Journal of Health Economics* 17(5):557-86.

[39] Parsons, T and Bales, R.F. 1955. *Family, Socialization and Interaction Process*. Glencoe, IL:Free Press.

[40] Pepler, D; Abramovitch, R and Corter, C. 1981. "Sibling Interaction in the Home: A Longitudinal Study" *Child Development* 52:1344-47.

[41] Powell, L.M., Tauras, J.A and Ross, H. 2005. "The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior" *Journal of Health Economics* 24(5):950-68.

[42] Price, Joseph. 2008. "Parent-Child Quality Time : Does Birth Order Matter?" *Journal of Human Resources* 43(1):240-65.

[43] Rosenberg, Florence R. and Robert.G. Simmons. 1975. "Sex Differences in the Self-Concept of Adolescence" *Sex Roles* 1(2):147-59.

[44] Staiger, Douglas and Stock, James H. 1997. "Instrumental Variables Regression with Weak Instruments " *Econometrica* 65(3):557-86.

[45] Steinberg, L. and Silverberg, S. 1986. "The Vicissitudes of Autonomy in Early Adolescence" *Child Development* 57: 841-51.

[46] Steinberg, L., Lamborn, S.D., Darling, N., Mounts, N.S and Dornbusch, S.M. 1994. "Overtime Changes in Adjustment and Competence among Adolescents from Authoritative, Authoritarian, Indulgent and Neglectful Families" *Child Development* 65: 754-70.

[47] Vidrine, J.I; Anderson, C.B; Pollak, K.I and Wetter, D.W. 2006. "Gender Differences in Adolescent Smoking: Mediator and Moderator Effects of Self-Generated Expected Smoking Outcomes" *American Journal of Health Promotion* 20(6):383-7.

[48] Widmer, Eric. 1997. "Influence of Older Siblings on Initiation of Sexual Intercourse" *Journal of Marriage and Family* 59(4):928-38.

[49] Windle, Michael. 2000. "Parental, Sibling and Peer Influences on Adolescent Substance Use and Alcohol Problems" *Applied Developmental Science* 4(2):98-110.

Table 1 : Descriptive Statistics from the NLSY97

A. Dependent Variables	**Mean**
# days smoked cigarettes in the last 30 days	
All respondents	5.56
Older Siblings	5.33
Younger Siblings	3.15
# days drunk alcohol in the last 30 days	
All respondents	1.8
Older Siblings	1.68
Younger Siblings	0.96

B. Explanatory Variables	**Mean (Smoking Sample)**	**Mean (Drinking Sample)**
Same Sex Pairs	0.51	0.51
Mixed Sex Pairs	0.49	0.49
Age		
Older Sibling	15.22	15.22
Younger Sibling	13.34	13.34
Live in an urban area	0.66	0.66
Mother's age at birth	25.78	25.77
White	0.69	0.69
Mother's Education		
HS dropout	0.15	0.15
HS graduate	0.34	0.34
Some college	0.24	0.24
College plus	0.27	0.27
Household Size	4.2	4.2
Imputed family income	$65862.65	$65768.28
Live with both parents	0.6	0.6
Hispanic	0.1	0.1
No. Of Observations	4227	4227

Table 2 : Descriptive Statistics from the NLSY97
Conditional on Smoking and Drinking

Explanatory Variables	**Smoke Cigarettes**		**Drink Alcohol**	
	No	**Yes**	**No**	**Yes**
Live in an Urban Area	0.64	0.69	0.66	0.66
Mother's age at birth	25.84	25.19	25.50	25.51
White	0.65	0.79	0.64	0.79
Mother's Education				
HS dropout	0.15	0.17	0.15	0.15
HS graduate	0.34	0.35	0.35	0.32
Some college	0.22	0.27	0.22	0.27
College plus	0.29	0.21	0.28	0.26
Household Size	4	3.6	4	3.5
Imputed family income	$66787.92	$57563.58	$65096.74	$58934.57
Live with both parents	0.68	0.58	0.66	0.59
Hispanic	0.12	0.08	0.11	0.10

Table 3 : Conditional Probabilities of the Younger Sibling's Substance Use

Smoke Cigarettes

	Older Sibling		
	No	Yes	Total
Younger Sibling			
No	2749	694	3443
	88.42%	62.07%	
Yes	360	424	784
	11.57%	37.92%	
Total	3109	1118	4227

χ^2 : 257.97
p-value : 0.000

Drink Alcohol

	Older Sibling		
	No	Yes	Total
Younger Sibling			
No	2343	887	3230
	84.34%	61.21%	
Yes	435	562	997
	15.65%	38.78%	
Total	2778	1449	4227

χ^2 : 175.56
p-value : 0.000

The percentages are column percentages, calculated on the basis of whether the older sibling smokes (drinks) or does not smoke (drink).

Table 4 : OLS Results for Smoking and Drinking

	(1) Days Smoked	(2) Days Smoked	(3) Days Drank	(4) Days Drank
Older Sibling's Behavior	0.27***	0.22***	0.10***	0.09**
	(0.018)	(0.021)	(0.020)	(0.047)
Male Young	-8.42***	-6.91***	-3.57***	-3.30***
	(1.430)	(1.973)	(0.577)	(0.834)
Age Young	0.21***	0.12**	0.06***	0.06***
	(0.021)	(0.067)	(0.007)	(0.019)
Male Young * Age Young	0.53***	0.43**	0.24***	0.22***
	(0.107)	(0.144)	(0.042)	(0.060)
Mother's Age at Birth	-	0.02	-	0.005
		(0.031)		(0.011)
White	-	2.13***	-	0.70***
		(0.351)		(0.117)
Live in Urban Area	-	-0.44	-	0.08
		(0.362)		(0.135)
Mother HS Dropout	-	2.66***	-	0.22
		(0.683)		(0.214)
Mother HS Graduate	-	1.52***	-	0.337*
		(0.4096)		(0.181)
Mother Some College	-	1.01***	-	-0.06
		(0.393)		(0.152)
Household Size	-	-0.19	-	-0.07
		(0.224)		(0.069)
Live with both parents	-	-2.74***	-	-0.52***
		(0.464)		(0.185)
Hispanic	-	-2.07	-	-0.13
		(0.457)		(0.191)
Elasticity	0.45	0.37	0.16	0.18
No. Of Observations	4227	4227	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income and time dummies.
Elasticities are evaluated at mean values of the outcome for older and younger children.

Table 5 : 2SLS Results for Smoking and Drinking

	(1) Days Smoked	(2) Days Drank
Older Sibling's Behavior	0.42***	0.17**
	(0.086)	(0.090)
Male Young	-5.16**	-3.14***
	(2.241)	(0.990)
Age Young	0.10	0.06***
	(0.068)	(0.021)
Male Young * Age Young	0.31**	0.21***
	(0.160)	(0.071)
Mother's age at birth	0.04	0.003
	(0.032)	(0.011)
White	1.06**	0.65***
	(0.568)	(0.175)
Live in Urban Area	-0.68*	0.08
	(0.374)	(0.133)
Mother HS Dropout	2.35***	0.22
	(0.704)	(0.214)
Mother HS Graduate	1.35***	0.32*
	(0.434)	(0.180)
Mother Some College	0.67	-0.07
	(0.430)	(0.150)
Household Size	-0.22	-0.07
	(0.222)	(0.074)
Live with both parents	-2.34***	-0.51***
	(0.491)	(0.180)
Hispanic	-1.39***	-0.11
	(0.535)	(0.200)
Elasticity	0.72	0.30
No. Of Observations	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income and time dummies.
Elasticities are evaluated at mean values of the outcome for older and younger children.

Table 7 : 2SLS Age Gap Results for Smoking and Drinking

	(1) Days Smoked	(2) Days Drank
Older Sibling's Behavior	0.98***	0.37***
	(0.109)	(0.128)
Older Sibling's Behavior * Big Age Gap	-0.47***	-0.20***
	-0.104	-0.070
Male Young	-3.91***	-1.76*
	(1.325)	(0.994)
Age Young	0.09	0.03
	(0.077)	(0.025)
Male Young * Age Young	0.12**	0.12*
	(0.058)	(0.075)
Mother's age at birth	0.07*	0.003
	(0.045)	(0.011)
White	1.55*	0.47***
	(0.960)	(0.192)
Live in Urban Area	-1.21***	0.14
	(0.494)	(0.134)
Mother HS Dropout	1.36**	0.21
	(0.752)	(0.215)
Mother HS Graduate	0.78	0.31*
	(0.575)	(0.181)
Mother Some College	0.37	-0.12
	(0.653)	(0.158)
Household Size	-0.33	-0.13
	(0.258)	(0.094)
Live with both parents	-1.33**	-0.47***
	(0.639)	(0.179)
Hispanic	-0.46	-0.07
	(0.818)	(0.200)
No. Of Observations	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income and time dummies.

Table 8 : 2SLS Results for the Smoking and Drinking Outcomes : By Sibship Gender Composition

	(1) Days Smoked	(2) Days Drank
Same Sex Older Sibling's Behavior	0.34***	0.15**
	(0.088)	(0.071)
Mixed Sex Older Sibling's Behavior	0.63***	0.33***
	(0.115)	(0.100)
Male Young	-5.13**	-2.86***
	(2.368)	(1.030)
Age Young	0.092	0.05***
	(0.069)	(0.021)
Male Young * Age Young	0.30*	0.20***
	(0.169)	(0.074)
Mother's age at birth	0.05*	0.007
	(0.034)	(0.016)
White	0.90**	0.57***
	(0.493)	(0.174)
Live in Urban Area	-0.54	0.13
	(0.391)	(0.135)
Mother HS Dropout	2.45***	0.26
	(0.720)	(0.213)
Mother HS Graduate	1.41***	0.35**
	(0.450)	(0.183)
Mother Some College	0.54	-0.05
	(0.449)	(0.150)
Household Size	-0.22	-0.05
	(0.231)	(0.075)
Live with both parents	-2.19***	-0.51***
	(0.504)	(0.180)
Hispanic	-1.15**	-0.07
	(0.538)	(0.198)
Same Sex Elasticity	0.62	0.28
Mixed Sex Elasticity	0.98	0.55
No. Of Observations	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income and time dummies.
Elasticities are evaluated at mean values of the outcome for older and younger children.

Table 9 : 2SLS Cross Substance Results for Smoking and Drinking

	(1) Days Smoked	(2) Days Drank
Older Sibling's Smoking	0.64**	0.05*
	(0.348)	(0.038)
Older Sibling's Drinking	0.20	0.18*
	(0.188)	(0.106)
Male Young	-4.17*	-3.13***
	(2.404)	(0.987)
Age Young	0.05	0.05**
	(0.084)	(0.025)
Male Young * Age Young	0.25***	0.21***
	(0.039)	(0.071)
Mother's age at birth	0.02	0.008
	(0.036)	(0.012)
White	1.43**	0.67***
	(0.642)	(0.190)
Live in Urban Area	-0.37	-0.09
	(0.477)	(0.132)
Mother HS Dropout	2.69***	0.12*
	(0.795)	(0.086)
Mother HS Graduate	1.51***	0.32*
	(0.469)	(0.197)
Mother Some College	0.98**	-0.15
	(0.506)	(0.157)
Household Size	-0.32	-0.06
	(0.240)	(0.0768)
Live with both parents	-2.64***	-0.52**
	(0.569)	(0.209)
Hispanic	-1.88***	-0.12
	(0.660)	(0.237)
No. Of Observations	4227	4227

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income and time dummies.

Table 10 : 2SLS Results for Smoking and Drinking : By Younger Child's Gender

	YOUNGER GIRLS		YOUNGER BOYS	
	(1)	**(2)**	**(3)**	**(4)**
	Days Smoked	**Days Drank**	**Days Smoked**	**Days Drank**
Older Sibling's Behavior	0.56***	0.38***	0.26**	0.05***
	(0.126)	(0.179)	(0.120)	(0.108)
No. Of Observations	2086	2086	2141	2141

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, time dummies, younger child characteristics and family factors.

Appendix 1 : Derivation of the Omitted Variable Bias

Consider the following estimated equation of the second child where *y* indexes the younger sibling and *o* indexes the older sibling,

$$Y_{yi} = \beta X_{yi} + \theta Y_{oi} + \epsilon_{yi}$$

Let us define the following vectors,

$$\begin{aligned} X_{yi} &= (1, age_{yi}, g_{yi}, age_{yi} * g_{yi}, Fam_i) \\ X_{oi} &= (age_{oi}, g_{oi}, age_{oi} * g_{oi}) \\ X_i &= (X_{oi}, X_{yi}) \end{aligned}$$

The above second stage regression is then estimated using instruments (X_{yi}, $\hat{Y_{oi}}$). However, since parental investments are omitted from the above regression and are not accounted for by the instruments, the true data generating process is,

$$Y_{yi} = \beta X_{yi} + \theta Y_{oi} + \gamma I_{yi} + \nu_{yi}$$

The asymptotic bias in $\hat{\theta}$ is then the coefficient on investments multiplied by the probability limit of the coefficient of $\hat{Y_{oi}}$ from an auxiliary regression of I_{yi} on the full set of included and excluded instruments. This can be written as,

$$\text{Asymptotic Bias} = \gamma \ \ \text{plim}\left[(\hat{Y_o}' M_y \hat{Y_o})^{-1})\hat{Y_o} M_y I_y\right]$$

where

$$M_y = I - X_y(X_y'X_y)^{-1}X_y$$

Now, let us specify the reduced form first stage equation and an auxiliary equation for parental investments in the younger child as follows,

$$Y_{oi} = \delta_1 X_{oi} + \delta_2 X_{yi} + \eta_1$$
$$I_{yi} = \pi_1 X_{oi} + \pi_2 X_{yi} + \eta_2$$

Using the relation $M_y X_y = 0$, the bias term reduces to the following expression,

$$\text{Asymptotic Bias} = \gamma \ \ (\delta_1' A \delta_1)^{-1} \delta_1' A \pi_1$$

where

$$A = \text{plim} \left[n^{-1} X_o' M_y X_o \right]$$

From the above, we can see that the condition for the asymptotic bias to vanish is $\delta_1' A \pi_1 = 0$. Since *A* and δ_1 cannot be equal to zero, the condition for the bias to disappear reduces to $\pi_1 = 0$. In effect, this implies that the correlation between $\hat{Y_o}$ and I_y should be zero after controlling for X_y.

Appendix 2

Table 11 : First Stage Results : Smoking

	(1) Pooled Sample	(2) Same Gender Sibs	(3) Mixed Gender Sibs	(4) Cross Substance (Drinking)
Male Old	-21.84***	-23.53***	-19.11***	-9.40***
	(3.181)	(3.683)	(4.497)	(1.203)
Age Old	0.66**	0.68**	1.28**	0.15**
	(0.278)	(0.324)	(0.388)	(0.080)
Male Old * Age Old	1.28***	1.23***	0.98	0.59***
	(0.198)	(0.458)	(0.566)	(0.077)
Mother's Age at Birth	0.09**	0.02	0.07**	0.02
	(0.043)	(0.034)	(0.038)	(0.015)
White	5.82***	3.46***	2.39***	1.44***
	(0.556)	(0.42)	(0.403)	(0.165)
Live in Urban Area	1.31***	1.43***	0.19	0.15
	(0.516)	(0.388)	(0.481)	(0.175)
Mother HS Dropout	1.85**	1.55**	1.27*	0.12
	(0.953)	(0.731)	(0.769)	(0.251)
Mother HS Graduate	1.01	1.32**	0.23	0.12
	(0.633)	(0.582)	(0.441)	(0.212)
Mother Some College	1.84***	0.75	1.144**	0.19
	(0.683)	(0.479)	(0.503)	(0.219)
Household Size	-0.02	-0.04	0.13	-0.1
	(0.289)	(0.24)	(0.185)	(0.093)
Live with both parents	-2.38***	-1.17***	-1.15***	-0.34*
	(0.621)	(0.470)	(0.429)	(0.184)
Hispanic	-3.63***	-1.63***	-1.82***	-0.49**
	(0.767)	(0.587)	(0.489)	(0.219)
No. Of Observations	4227	4227	4227	4227
Staiger and Stock First Stage F-Statistic	23.65	36.69	28.18	23.06
P-value of Excluded Instruments	0.000	0.000	0.000	0.000

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income, time dummies and the younger sibling's age and gender.

Table 12 : First Stage Results : Drinking

	(1) Pooled Sample	(2) Same Gender Sibs	(3) Mixed Gender Sibs	(4) Cross Substance (Smoking)
Male Old	-9.41***	-13.81***	-8.71***	-21.92***
	(1.203)	(1.600)	(1.538)	(3.182)
Age Old	0.14*	0.11*	0.46***	0.67***
	(0.080)	(0.063)	(0.076)	(0.278)
Male Old * Age Old	0.59***	0.85***	0.32**	1.28***
	(0.077)	(0.166)	(0.172)	(0.199)
Mother's Age at Birth	0.02	0.02	0.06	0.09**
	(0.015)	(0.012)	(0.085)	(0.043)
White	1.44***	0.76***	0.68***	5.84***
	(0.165)	(0.104)	(0.134)	(0.555)
Live in Urban Area	0.15	0.20*	-0.21*	1.31***
	(0.175)	(0.116)	(0.131)	(0.516)
Mother HS Dropout	0.07	0.32**	-0.18	1.85**
	(0.249)	(0.165)	(0.183)	(0.956)
Mother HS Graduate	0.11	0.17	-0.07	1.03
	(0.212)	(0.147)	(0.154)	(0.633)
Mother Some College	0.18	0.24	0.06	1.92***
	(0.219)	(0.152)	(0.159)	(0.684)
Household Size	-0.10	-0.05	-0.02	-0.04
	(0.093)	(0.068)	(0.062)	(0.289)
Live with both parents	-0.349**	-0.28**	-0.14**	-2.38***
	(0.183)	(0.140)	(0.065)	(0.621)
Hispanic	-0.52**	-0.14	-0.31**	3.65***
	(0.218)	(0.174)	(0.136)	(0.768)
No. Of Observations	4227	4227	4227	4227
Staiger and Stock First Stage F-Statistic	22.94	35.23	34.82	23.69
P-value of Excluded Instruments	0.000	0.000	0.000	0.000

Robust standard errors in brackets clustered at sibling pair.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, family income, time dummies and the younger sibling's age and gender.

Table 13 : Exogeneity of Instruments with respect to Investments

Investment Measure	**Chi Square Statistic**
Mother helps child with what is important	2.02
Mother knows who child is with when not at home	1.74
Mother knows child's teachers and school activities	10.66*
Mother knows child's close friends	1.22
Degree of monitoring by mother	3.14
Degree of child relationship with mother	5.32
Father helps child with what is important	4.18
Father knows who child is with when not at home	2.43
Father knows child's teachers and school activities	4.17
Father knows child's close friends	2.35
Degree of monitoring by father	6.39
Degree of child relationship with father	5.6
Total allowance received in the past year	0.27

the Chi-square statistic is from a joint test of the excluded instruments.
*, ** and *** denote significance at 1%, 5% and 10% levels respectively.
All regressions include a constant, time dummies, younger child characteristics and family factors.