Is There a Stepping Stone Effect in Drug Use? Separating State Dependence from

Unobserved Heterogeneity Within and *Between* Illicit Drugs

Monica Deza[1]

University of Texas at Dallas

Abstract:

Empirically, teenagers who use alcohol or marijuana in one period are more likely to use hard drugs in the future. This pattern can be explained by a causal effect of soft drug consumption on future consumption of hard drugs (i.e., state dependence *between* drugs or stepping-stone effects) or by unobserved characteristics that make people more likely to use both soft and hard drugs (i.e., correlated unobserved heterogeneity). I estimate a dynamic discrete choice model of alcohol, marijuana and hard drug use over multiple years, and separately identify the contributions of state dependence and unobserved heterogeneity. I find modest-sized but statistically significant "stepping-stone" effects from softer to harder drugs that are largest among the youngest individuals in my sample. This study also suggests that alcohol, marijuana and hard drugs are complements in utility.

JEL: I10, C10, C33.

Keywords: state dependence, stepping-stone, illicit drugs.

---

[1] Corresponding author.
Email address: monica.deza@utdallas.edu
Address: 800 West Campbell Road GR31, Richardson, TX 75080
Fax number: (972)883-6486

I.INTRODUCTION

Heated debates have arisen as states such as Colorado and Washington have decriminalized marijuana at the same time as the federal government has continued enforcement of laws against the drug. Given the limited evidence on the health impacts of marijuana (e.g., Prinz 1997), supporters of the federal position have often implicitly relied on the argument that use of marijuana leads to an increased use of harder and more socially disruptive drugs such as cocaine and amphetamines. Casual observation suggests that most users of hard drugs start off using alcohol and/or marijuana. Whether the use of softer drugs actually *causes* the future use of hard drugs (or the continued use of soft drugs) is unclear. The dynamic patterns could arise from a change in preferences that occurs among those who use softer drugs, i.e., a true "state dependence" effect (Heckman, 1981a). Alternatively, they could simply reflect the fact that certain individuals are more likely to consume drugs at any point in time -- a heterogeneity effect.

Disentangling the stepping-stone effect from unobserved heterogeneity is potentially important for policy. For instance, assume that a policy maker's goal is to reduce long term cocaine use. If there is a stepping-stone effect from marijuana to cocaine, any small shock that leads some young people to use marijuana at some point in time will have a long-term effect on further use of harder drugs. Consequently, policies to prevent marijuana use can be an effective channel for preventing long term cocaine use. Similarly, if there is state dependence in cocaine use, policies that prevent cocaine consumption at younger ages may have a lasting benefit in reducing longer-term cocaine use

Statistical models that separate state dependence from unobserved heterogeneity have been widely used to model welfare participation (Plant, 1984; Engberg, Gottschalk and Wolf, 1990; Card and Hyslop 2005), dynamic labor supply of married women (Hyslop 1999), self-

reported health (Halliday, 2008), sexual behavior among teenagers (Arcidiacono, Khwaja, Ouyang, 2009), charitable giving (Meer, forthcoming) and many other outcomes. These models are also used in marketing to separate tastes based on habit formation in brand purchases (e.g., Keane 1997). In the marketing literature, the estimated parameters are often used to simulate the effects of a shock on consumption of a particular brand (e.g., caused by a promotion or sale) on the long term purchases of the brand. My goal in this paper is similar. In particular, I use the estimated model of dynamic drug use to simulate whether an exogenous shock that reduces marijuana or alcohol consumption will have a long term effect on cocaine use. I develop a series of multiple-equation logit and probit style models with unobserved heterogeneity and state dependence that allow me to estimate "within-drug" state dependence (e.g., the effect of current alcohol use on future use) and "between-drug" state dependence or stepping stone effects[2] (e.g., the effect of current alcohol use on future cocaine use). I consider models with first order state dependence, as well as models with higher order dependence and with heterogeneous state dependence. In addition, I consider "ordered" models that allow me to distinguish between different levels of intensity of drug use at each point in time. Finally, I consider models that separate state dependence within and between drugs from complementarity. Throughout, I use mass-point mixing models to non-parametrically account for time-invariant multidimensional unobserved heterogeneity (Heckman and Singer, 1984).[3] My models also include flexible

---

[2] Previous research uses the term "stepping-stone effect" from drug k to drug j to describe how initiation into drug k increases the likelihood of initiation into a harder drug j. For this paper, I refer to the stepping stone effect from drug k to drug j as the causal effect of current drug k consumption on future drug j consumption. State dependence between drug k and j comprises the stepping stone effect from drug k to drug j and the reverse stepping stone effect from drug j to k, where k is a softer drug than drug j (alcohol is softer than marijuana and marijuana is softer than cocaine). Stepping stone effects are analogous to state dependence between two drugs.

[3] That is, I let the data tell me whether individuals who have a high time-invariant preference for marijuana also have a high time-invariant preference for cocaine and alcohol. I treat the distribution of the unobserved component as discrete and drawn from the mixture distribution (Heckman, Singer, 1984). Each type m is assigned a vector $(\alpha_m^{drink}, \alpha_m^{mar}, \alpha_m^{coc}, \alpha_m^{drink_{ic}}, \alpha_m^{mar_{ic}}, \alpha_m^{coc_{ic}})$. This specification allows correlation between these six random effects within type.

controls for the initial conditions problem caused by the fact that some individuals in my data source, the National Longitudinal Survey of Youth 1997 (NLSY97), are first interviewed after they have already initiated soft (or even hard) drug use (Heckman 1981b; Wooldridge 2005).[4]

Much of the existing drug-use literature has overlooked the role of individual preferences in drug consumption, and interpreted the fact that most young adults consume marijuana before consuming cocaine as evidence of a "gateway" effect (Mills and Noyes 1984, Newcomb and Bentler (1986), Kandel and Yamaguchi (1984) among many others). A notable exception is Van Ours (2003), who uses a mixed proportional hazards model to study the extent to which first-time marijuana consumption affects first-time cocaine consumption. Van Ours (2003) concludes that, while marijuana initiation has a significant stepping stone effect on the future initiation of cocaine, the main factor driving the initiation of both drugs is unobserved heterogeneity.

Relative to the existing literature, I make two main contributions. First, I extend the consideration of state dependence and unobserved heterogeneity to a multiproduct setting, where the outcomes cannot be bundled into mutually exclusive classes. Second, to the best of my knowledge, I am the first to consider stepping-stone effects in a general dynamic setting where past use of each of several drugs can affect the decision to use each drug today. Looking at the effect of each of the three drugs on future consumption patterns allows me to compare the relative size of the stepping-stone effects of marijuana and alcohol on cocaine use. I can also test whether cocaine use causes increased future consumption of softer drugs (a "reverse" stepping-stone effect).

My empirical results suggest that softer drugs have a stepping-stone effect on harder drugs that is highly robust across specifications. That is, alcohol use has a positive stepping-stone effect on future use of marijuana and cocaine, and marijuana use has a positive stepping-stone

[4] NLSY97 started collecting data on cocaine use starting in 1998.

4

effect on future use of cocaine. The "reverse" stepping stone effect from harder to softer drugs is statistically insignificant in most of my specifications, and is uniformly smaller than the effect from softer to harder drugs, indicating that the primary stepping-stone effect operates from softer to harder drugs. The estimated stepping-stone effect from alcohol to cocaine is comparable in size to the effect from marijuana to cocaine, suggesting that policy concerns about stepping stone effects should include both substances.

I also find strong evidence that both the permanent and transitory unobservable components of tastes are correlated across drugs. In addition, I extend previous models that distinguish correlation in preferences from the role of drugs as complements or substitutes (Getzkow, 2007) to also include state dependence within and between drugs. I find that drugs are complements in utility. For instance, consuming alcohol and marijuana together leads to higher utility than the sum of the utility derived from consuming alcohol alone and marijuana alone. However, the state dependence (within and between drugs) parameters remain robust to the inclusion of complementarity.

Furthermore, the stepping-stone effect is heterogeneous across people. In particular, the stepping-stone effect is largest among young people, indicating that early consumption of softer drugs may have an impact on consumption of harder drugs. On the other hand, state dependence for each of the three drugs increases with age, indicating that the habit of consuming a particular drug may be harder to break with age. The stepping-stone effect from softer to harder drugs is greater among those who use the relatively softer drug more heavily, and state dependence for each drug is also higher among heavy users. Most of the persistence of drug use for a particular drug is driven by state dependence, while the stepping-stone effect plays a minor role in

explaining why the probability of consuming harder drugs is higher among those who consumed softer drugs in the previous period.

These results are robust to the inclusion of demographics (race, gender, single parent household) and time-varying variables (region of residence, enrollment in a two-year or four-year college, employment status, and state-level current unemployment rate). Finally, these results also remain robust to the inclusion of drug and time specific covariates such as state-level share of drug treatment admissions that were attributed to alcohol, marijuana or hard drugs, as well as prices.

A limitation of the models I develop in this paper is the assumption that unobserved heterogeneity in tastes for drug use can be decomposed into the sum of a purely permanent component and a purely transitory component. To examine the extent to which this assumption causes misspecification, I present a variety of specification diagnostics, including a sample-analogue of generalized residuals that allows me to test whether there is evidence of serial correlation in use patterns after accounting for permanent and transitory taste shocks. I also use sample-analogue generalized residuals to diagnose misspecifications arising from contemporaneous correlations in the transitory taste components driving consumption of the three drugs.

This paper is organized as follows. The next section discusses the data, while section 3 presents models with homogeneous state dependence and stepping-stone effects. Section 4 discusses empirical results, evaluates goodness of fit and presents specifications diagnostics Section 5 presents models with heterogeneous state dependence and stepping-stone effects, models with time and drug specific exclusion restrictions, and models that separate

complementarity from correlation in preferences, and presents counterfactual experiments.

Finally, section 6 summarizes and concludes.

II.DATA

I use restricted data from the National Longitudinal Survey of Youth 1997. The NLSY97 survey collects longitudinal information for a sample of 8,984 adolescents who were between the ages of 12 and 18 in 1997 (and between 22 and 28 in 2007).[5] Given that the main goal of this study is to model dynamic patterns of drug consumption over the respondents' life course, the NLSY97 is a nearly ideal data set. In each wave, participants report consumption of alcohol, marijuana, and cocaine in the last year, as well as frequency and intensity of consumption.[6] Measures of alcohol and marijuana use are available starting in 1997, while measures of cocaine use are available from 1998 onwards.[7]

I use data from the Treatment Episode Data Set-Admissions (TEDS-A) to compute drug-year and state specific share of drug treatment admissions that were attributed to alcohol, marijuana or hard drugs. The TEDS-A is an administrative dataset that provides information about the state-level flow of admissions to providers of substance abuse treatment[8]. The TEDS-A series provides individual level data for persons who were admitted to public and private substance abuse treatment programs receiving public funding. This data includes a variable on year and state in which the admission to treatment took place, as well as the substances for which

---

[5] The sample is based on a stratified design and includes an "oversample" of minorities. In this paper I make use of the entire sample, and make no allowance for sample weights.

[6] The NLSY97 asks respondents "Excluding marijuana and alcohol, since the date of the last interview, have you used any drugs like cocaine or crack or heroin, or any other substance not prescribed by a doctor, in order to get high or to achieve an altered state?" While this measure of hard drug use includes cocaine and other hard drugs, I will refer to it as cocaine use for the remaining of the paper.

[7] The respondents were asked the following questions regarding cocaine in 1998: (1)"Have you ever used cocaine?" and (2) "Number of times the respondent used cocaine/hard drugs since the date of last interview." I create an indicator for whether the respondent consumed cocaine in the last year, where I assign a 1 if the respondent reported having used cocaine at least once since the last interview and a 0 otherwise

[8] This dataset corresponds to ICPSR study 25221. I acknowledge that SAMHSA and ICPSR will bear no responsibility for my use of the data or for my interpretations or inferences based upon such use.

the individual was admitted into treatment. A limitation of this dataset is that TEDS-A includes only admissions to facilities that are licensed or certified by the state substance abuse agency or that are administratively tracked for other reasons. In addition, there may be variations in state reporting practices. However, "TEDS was estimated to include 83 percent of TEDS-eligible admissions and 67 percent of all known admissions" in 1997. [9]

To measure year and region specific prices of alcohol, I use the Average Price Index for malt (by region and month) provided by the Bureau of Labor Statistics (BLS).[10] I compute a yearly average and convert this to 1980-1984 dollars.

Finally, to measure drug and time specific prices of marijuana and hard drugs, I use data from the "Price and Purity of Illicit Drugs: 1981-2007" report. This report provides prices and measures of purity of selected drugs such as powder cocaine, crack cocaine, heroin, d-methamphetamine, and marijuana. The estimates in this report are based on data from System to Retrieve Information from Drug Evidence (STRIDE) database maintained by the Drug Enforcement Administration (DEA).[11] I convert these prices to 1980-1984 dollars.

I will use drug-year and state specific share of drug treatment admissions, price per expected pure gram of hard drugs, and price per bulk gram of marijuana as exclusion restrictions. I will describe in more detail the limitations of each.

---

[9] Source: http://www.icpsr.umich.edu/icpsrweb/SAMHDA/series/56

[10] I use nominal malt prices by region (urban) and year for malt beverages of all types per 16 oz. or 473 ml. This data is reported by month, and I calculate the yearly average. The BLS sequences are as follows: Midwest APU0200720111, Northeast APU0100720111, South APU0300720111 and West APU0400720111.

[11] Office of National Drug Control Policy (2004). *The Price and Purity of Illicit Drugs: 1981 Through the Second Quarter of 2003*, Washington, D.C.: Executive Office of the President (Publication Number NCJ 207768), electronically accessible through the following World Wide Web address http://www.whitehousedrugpolicy.gov/publications/price_purity/. The accompanying *Technical Report* is available at http://www.whitehousedrugpolicy.gov/publications/price_purity_tech_rpt/. Sponsored by ONDCP, both reports were produced at RAND's Drug Policy Research Center and Public Safety and Justice Division.

Table 1 presents an overview of the characteristics of the sample from NLSY97. Column 1 shows descriptive data for the entire sample. Column 2 summarizes my main analysis sample, which is restricted to individuals who were not lost due to attrition from 1997 to 2007, and who did not avoid answering those questions. Columns 3, 4 and 5 show characteristics for subsets of respondents in the analysis sample who consumed alcohol, marijuana, or cocaine in at least one year during my sample period.

The difference across subsamples in Panel A can be summarized as follows. First, most of the demographic characteristics of the entire sample are statistically indistinguishable from those of the subsample that was not lost due to attrition after ten waves, though men are more prone to attrition than women.[12] Second, alcohol use is nearly universal (95% have ever used alcohol) while marijuana is highly prevalent (57%) and cocaine[13] use is less so (25%). Third, as might be expected, the average starting age for alcohol is lower than the starting age for marijuana, which is lower than the starting age for cocaine. Fourth, individuals who ever use marijuana have a very high rate of ever using alcohol (99%) and a lower starting age for alcohol. Likewise, 100% of ever-users of cocaine have consumed alcohol, and 93% have consumed marijuana, and this group has the lowest starting ages for alcohol and marijuana among the subgroups in the table.[14] The data on the order of the first use of the three substances, shown in Panel B of Table 1, are consistent with these patterns. Among respondents who eventually consume all three drugs, most of them follow the sequence of first alcohol, second marijuana, and third cocaine. Moreover, use of cocaine before either marijuana or alcohol is very rare.

---

[12] I control for gender in the empirical part of this paper.

[13] The NLSY97 asks respondents whether they have consumed "any drugs like cocaine or crack or heroin, or any other substance not prescribed by a doctor, in order to get high or to achieve an altered state." For the remaining of this paper, I will refer to this category of hard drugs as cocaine.

[14] These percentages are higher among respondents who consumed other drugs. While 24.63% of the respondents who were not lost due to attrition consumed cocaine at least once, this percentage was 39.78 among those who consumed marijuana at least once.

Appendix B discusses the extent to which misreporting, attrition, and avoidance of drug-related questions affect the estimated parameters. Approximately, 94.59%, 92.78% and 94.47% of respondents who were not lost due to attrition reported non-missing answers to the alcohol, marijuana, and cocaine-related questions for all 10 waves, respectively. Furthermore, the NLSY97 collects answers to sensitive questions using audio computer-assisted self-interview (ACASI) which is associated with less underreporting of risky behaviors than other interview methods (Brener et al. 2003). Finally, appendix B discusses a simulation exercise to quantify the extent to which misclassification error, if present, affects the estimated parameters. Appendix C compares the NLSY97 to two other major datasets on drug use: National Study of Drug Use and Health (NSDUH), and Monitoring the Future (MTF).

More information on the dynamic patterns of drug use in my sample is contained in Figures 1 and 2, which illustrate the two key stylized facts that motivate my analysis. First, Figure 1 shows that the probability of consuming a particular drug at any given period is higher among those who consumed that drug in the previous period. This pattern holds for alcohol, marijuana, and cocaine[15]. Second, Figure 2 illustrates that the probability of consuming cocaine is higher among those who consumed alcohol or marijuana in the previous period than among those who abstained from using these drugs. Similarly, the probability of consuming marijuana is higher among those who consumed alcohol in the previous period. These patterns are exhibited in Figure 1 and 2.The key econometric goal of this paper is to explain the degree to which these patterns occur because of true state dependence within and between drugs (stepping-stone effects), instead of correlated preferences.

---

[15] For instance, the probability of consuming marijuana at any period among those who consumed it in the previous period is 66.99%. On the other hand, among those who abstained from using marijuana in the previous period, this probability is 9.09%. A similar pattern holds for alcohol and cocaine.

III MODEL, IDENTIFICATION AND ESTIMATION

First, I experiment with three models with the objective of disentangling true state dependence and stepping-stone effects from unobserved heterogeneity. To keep these models as simple as possible, I begin by assuming homogeneous first order state dependence and stepping-stone effects. I first consider a simple trivariate logit model with only first-order state dependence (Model A), then consider a trivariate probit model (Model B), which allows correlation across the transitory preference shocks in each period. Then I consider a trivariate logit model with second-order state dependence (Model C).

A. Model A: TRIVARIATE LOGIT MODEL with first-order State Dependence

Consider an individual who maximizes his or her utility by choosing whether to consume three non-mutually exclusive drugs: alcohol, marijuana, and cocaine[16]. Let U_{ijt} denote the utility that individual i experiences from consuming drug j in year t, and let $Y_{i,k,t}$ represent an indicator that equals 1 if person i consumed drug k in period t. I assume that the utility of drug use in period t depends on drug-specific time trend[17] $\delta_{0j} + \delta_{1j}(t - t_0)$, on a set of observed characteristics (X) of the individual, [18] on lagged indicators for drug use in period t-1, and on a combination of a permanent unobserved taste component α_{ij} and a transitory component ε_{ijt}:

$$U_{ijt} = \underbrace{\delta_{0j} + \delta_{1j}(t - t_0) + X_{it}\beta_j + \sum_{k=1}^{J} \gamma_{kj}Y_{i,k,t-1} + \alpha_{ij}}_{V_{ijt}(\alpha_{ij})} + \varepsilon_{ijt} \qquad (1)$$

---

[16] In a previous version of the paper, I also estimate these models using tobacco, marijuana and cocaine. A logit model estimates that tobacco has a smaller effect than alcohol on future cocaine use, and that this effect becomes negligible when I allow for transitory shocks across drugs to be correlated (probit specification). Also, the NLSY97 groups cocaine with other hard drugs.

[17] The variable t is the year that corresponds to U_{ijt}, while t_0 is always 1998.

[18] The vector X_{it} represents time-invariant characteristics (gender and whether the individual comes from a single-headed household) and time-varying observable characteristics (age) of consumer i in time t. Appendix D discusses models with different sets of demographics (race, ethnicity, whether the individual was in an urban or rural area in the first wave, and region of residence in the first wave), and time-varying covariates (whether the respondent is currently enrolled in a two-year college, currently enrolled in a four-year college, or currently employed, and finally the state-level unemployment rate in year t). The results remain robust to the inclusion of these covariates.

I assume that the non-random part of the utility $(V_{ijt}(\alpha_{ij}))$ derived from not using drug j in period t is zero. To complete the model, I need a specification for the outcomes of the "initial conditions," which are drug use outcomes in the initial year of my sample (1998).[19] I assume that the initial choices are selected according to the utility:

$$U_{ij0} = \underbrace{\omega_{0j} + X_{i0}\lambda_{0j} + \alpha_{ij0}}_{V_{ij0}(\alpha_{ij})} + \varepsilon_{ijt}^{initial} \tag{2}$$

where again the non-random part of the utility $(V_{ij0}(\alpha_{ij}))$ of not consuming drug j in year 1998 (period 0) is set to 0. In equation (1), the state dependence (within and between drugs) parameters are represented by the parameters γ_{kj}. For k=j these are the effects of lagged consumption of drug k on current consumption of drug k (within drug state dependence). For k≠j these are the stepping stone effects from use of drug k in the past to taste for drug j in the current period (state dependence between drug k and j). The parameters α_{ij} represent the time-invariant unobserved tastes of person i for drug j, which I assume are distributed across the population with a discrete distribution with a relatively small number of points of support (up to 7).[20] Finally, ε_{ijt} and $\varepsilon_{ijt}^{initial}$ represent transitory taste shocks, which in the logistic model are assumed to be drawn independently from an Extreme Value Type 1 distribution.

At each time period t, individual i chooses to use drug j if the utility derived from doing so is higher than the utility attained from not consuming drug j. With a logistic error assumption,

---

[19] A separate equation for initial conditions for each drug is required for two reasons. First, we do not observe data since the year in which the stochastic process started, and thus I cannot construct the likelihood function for all years in which the data generating process has been in operation (Keane 1997, Heckman 1981b, Wooldridge 2005). Given that there is serial correlation, ignoring the initial conditions problem leads to biased and inconsistent parameter estimates (Heckman, 1981b). Second, a separate equation is also needed because the initial period does not have lagged values and requires a different specification. Studies where the respondents have identical outcomes in the pre-sample periods do not face the initial conditions problem (Card and Hyslop 2005), while studies where the stochastic process of the relevant outcome started prior to the observed periods handle the initial conditions problem the same way I do in this study (Altonji et al 2010).

[20] For Model A, I experiment with 3,4,5,6, and 7 mass points, but I only report the estimates for the model with seven mass points, since allowing for a seventh mass point resulted in a significant improvement in the fit of the model, as measured by the log-likelihood, Pearson-Goodness-of –Fit, Akaike and Bayesian Information Criterion.

the probability of an individual using or abstaining from using drug j at any given period has a

closed form solution and can be written as equations (3a) and (3b) respectively:[21]

$$P\big(Y_{ijt} = 1\big|\bar{Y}_{i,t-1}, \alpha_{ij}\big) = \frac{\exp\left(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij}\right)}{1+\exp\left(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij}\right)} \tag{3a}$$

$$P\big(Y_{ijt} = 0\big|\bar{Y}_{i,t-1}, \alpha_{ij}\big) = \frac{1}{1+\exp\left(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij}\right)} \tag{3b}$$

Similarly, the probability of an individual using or abstaining from using drug j at the

initial period (1998) can be written as in equation (4a) and (4b) respectively

$$P\big(Y_{ij0} = 1\big|\alpha_{ij0}\big) = \frac{\exp\left(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0}\right)}{1+\exp\left(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0}\right)} \tag{4a}$$

$$P\big(Y_{ij0} = 0\big|\alpha_{ij0}\big) = \frac{1}{1+\exp\left(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0}\right)} \tag{4b}$$

I compute the likelihood of a sequence of indicators of drug j consumption $L_i^j(Y_{ij0,...,}Y_{ijT})$

by taking the weighted average of type-specific likelihood contributions, using the unconditional

probabilities, π_m, as weights.

$$L_i^j\big(Y_{ij0},..Y_{ijT}\big) = \sum_{m=1}^{M}\pi_m\{P(Y_{ij0}|\alpha_{mj0})\prod_{t=1999}^{T=2007}P(Y_{ijt}|\bar{Y}_{i,t-1},\alpha_{mj})\} \tag{5}$$

The individual contribution to the entire likelihood function can be written as

$$L_i = L_i^{alcohol}L_i^{marijuana}L_i^{cocaine} \tag{6}$$

With 7 mass points there are 87 parameters to estimate: nine utility parameters for each

drug j ($\delta_{0j},\delta_{1j},\beta_1,\beta_2,\beta_3,\beta_4,\gamma_{1j},\gamma_{2j},\gamma_{3j}$), 13 unobserved heterogeneity parameters for each drug

---

[21] $\bar{Y}_{i,t-1} = (Y_{i,t-1}^{drink}, Y_{i,t-1}^{mar}, Y_{i,t-1}^{coc})$, where $Y_{i,t-1}^{drink}, Y_{i,t-1}^{mar}, Y_{i,t-1}^{coc}$ represent the first-order lagged indicators of use $(Y_{i,j,t-1})$ of alcohol, marijuana and cocaine respectively. For the remainder of the paper, the following probabilities $P\big(Y_{ijt} = 1\big|\bar{Y}_{i,t-1}, \alpha_{ij}\big), P\big(Y_{ijt} = 0\big|\bar{Y}_{i,t-1}, \alpha_{ij}\big), P(Y_{ij0} = 1|\alpha_{ij0}), P(Y_{ij0} = 0|\alpha_{ij0})$ also condition on X_{it} implicitly, to save on notation.

$(\alpha_2^j, ..., \alpha_7^j, \alpha_{1,0}^j, ..., \alpha_{7,0}^j)^{22}$, 5 initial conditions parameters for each drug $(\omega_{0j}, \lambda^1_{0j}, \lambda^2_{0j}, \lambda^3_{0j}, \lambda^4_{0j})$,

and 6 type-associated probability parameters $(\phi_2, ..., \phi_7)$.[23]

B. Model B: TRIVARIATE PROBIT WITH FIRST-ORDER STATE DEPENDENCE

While Model A provides a guide to the potential stepping-stone effects, it is oversimplified, as it requires assuming that the transitory shocks are independent over time and across drugs. More realistically, the transitory shocks to demand for one drug may be correlated with the transitory shocks in demand for the others (e.g., moving to a drug-friendly college town is associated with a positive shock in demand for alcohol, marijuana and cocaine). Model B relaxes the assumption of uncorrelated transitory shocks by allowing the $\left(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right)$ to be distributed as trivariate normals, with an arbitrary correlation structure.[24] This model still imposes the restriction that there is no serial correlation in the transitory shocks either within or between drugs.

In particular, for Model B I also assume that ε_{ijt} has the following multivariate normal distribution,

$$\left(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & p_{12} & p_{13} \\ p_{12} & 1 & p_{23} \\ p_{13} & p_{23} & 1 \end{pmatrix}\right) \tag{7}$$

---

[22] I normalize the random effect for type m=1 for each drug j, α_1^j, to zero.

[23] I also normalize ϕ_1 to zero where ϕ_m enters the unconditional probability of being type m in a model with three types, written as follows

$$\pi_m = \frac{\exp(\phi_m)}{1 + \exp(\phi_2) + \exp(\phi_3)}$$

[24] To establish complementarity or substitutability across products, in the sense that the utility of consuming both products together is higher than consuming them separately, I would need higher-frequency data. Because the NLSY97 reports yearly data, I can only observe whether individuals consumed a combination of drugs in the same year. Annual data only allows me to establish correlation of the time-varying unobservable component across drugs. While such correlation will partially capture whether those drugs are true complements or substitute (in the sense that the utility of consuming some drugs together may be higher than consuming them separately), I will not be able to disentangle true complementarity or substitutability from mere correlation between the time-varying unobservable components across drugs.

As opposed to the logistic model, the correlated normal model requires that I estimate the

likelihood for the full set of indicators of drug use in each period, $L_i(\bar{Y}_{i0,...,}\bar{Y}_{iT})$, where

$$\bar{Y}_{it} = (Y_{it}^{drink}, Y_{it}^{mar}, Y_{it}^{coc})$$,

$$L_i(\bar{Y}_{i0}, ... \bar{Y}_{iT}) = \sum_{m=1}^{M} \pi_m \{P(\bar{Y}_{i0}|\alpha_{m,0}) \prod_{t=1999}^{T=2007} P(\bar{Y}_{it}|\bar{Y}_{i,t-1}, \alpha_m)\} \qquad (8)$$

The probabilities of a particular consumption bundle $\overline{Y_{it}}$, $P(\overline{Y_{it}} \mid \alpha_m)$ no longer have a

closed form solution under the trivariate probit specification. In fact, it is the cumulative

distribution function of a joint normal distribution, which is numerically approximated by a triple

integral.[25] In addition to the parameters estimated under the logistic distribution, the multivariate

probit model also estimates the correlation coefficients $p_{12} = cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar})$, $p_{13} =$

$cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{coc})$, and $p_{23} = cov(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc})$.

C.Model C: TRIVARIATE LOGIT WITH SECOND-ORDER STATE DEPENDENCE

To improve upon the previous specifications, which only allowed lagged drug

consumption from the previous year to affect current consumption, I estimate a multivariate logit

model that includes second-order state dependence and an interacted term of first and second-

order state dependence. While a model with second-order state dependence still requires

assuming that transitory shocks are independent over time, incorporating a second-order lagged

outcome decreases the extent to which first and second-order serial correlations are problematic.

I do not include second-order stepping-stone effects because section 6 indicates that correlation

---

[25] For instance,

$$P(\bar{Y}_{it} = (1,1,0)|\bar{Y}_{i,t-1}, \alpha_m) = P(-\varepsilon_{it}^{drink} < V_{it}^{drink}(\alpha_m), -\varepsilon_{it}^{mar} < V_{it}^{mar}(\alpha_m), \varepsilon_{it}^{coc} < -V_{it}^{coc}(\alpha_m))$$
$$= \Phi(V_{it}^{drink}(\alpha_m), V_{it}^{mar}(\alpha_m), -V_{it}^{coc}(\alpha_m); p_{12}, -p_{13}, -p_{23})$$

where Φ represents the trivariate normal cumulative distribution function of $(-\varepsilon_{it}^{drink}, -\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc})$.

of transitory shocks across drugs and periods (assumption 3 from Model A) is not a source of misspecification in the benchmark model A.

An issue for a second order model is the specification of the initial conditions. These consist of the drug use choices in the two initial years 1998 and 1999: for each drug there are four mutually exclusive potential outcomes (0,0), (1,0),(0,1), and (1,1). The first component indicates drug consumption in 1998 and the second one indicates this for 1999.[26] Each initial conditions equation has a separate location parameter.[27] The latent utilities for each drug j are as follows:

$$U_{ijt} = \underbrace{\delta_{0j} + \delta_{1j}(t - t_0) + X_{it}\beta_j + \sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1} + \lambda_{0j}Y_{i,j,t-2} + \lambda_{1j}Y_{i,j,t-1}Y_{i,j,t-2} + \alpha_{ij}}_{V_{ijt}(\alpha_{ij})} + \varepsilon_{ijt} \quad (9)$$

$$U_{ij0}(0,0) = \varepsilon_{ij0}^{(0,0)} \quad (10)$$

$$U_{ij0}(0,1) = \underbrace{X_{i0}\omega_{0j}^{(0,1)} + \alpha_{ij0}^{(0,1)}}_{V_{ij0}^{(0,1)}} + \varepsilon_{ij0}^{(0,1)} \quad (11)$$

$$U_{ij0}(1,0) = \underbrace{X_{i0}\omega_{0j}^{(1,0)} + \alpha_{ij0}^{(1,0)}}_{V_{ij0}^{(1,0)}} + \varepsilon_{ij0}^{(1,0)} \quad (12)$$

$$U_{ij0}(1,1) = \underbrace{X_{i0}\omega_{0j}^{(1,1)} + \alpha_{ij0}^{(1,1)}}_{V_{ij0}^{(1,1)}} + \varepsilon_{ij0}^{(1,1)} \quad (13)$$

The multinomial logit specification assumes that $\varepsilon_{ij0}^{(0,0)}$, $\varepsilon_{ij0}^{(0,1)}$, $\varepsilon_{ij0}^{(1,0)}$, and $\varepsilon_{ij0}^{(1,1)}$ are independent. This assumption is not as restrictive as it might first seem, given that I have modeled the initial conditions separately, allowing for a different random effect at each potential initial outcome.

---

[26] While I model the initial conditions with multinomial logit because there are four mutually exclusive outcomes for each drug, the entire model is still a multivariate or generalized logit model with three non-mutually exclusive binary outcomes.

[27] The model estimates the specified distribution of the unobserved heterogeneity with M discrete points of support, where each point of support m corresponds to the following vector of unobserved heterogeneity:
$(\alpha_{m,}^{drink}, \alpha_{m,}^{mar}, \alpha_{m,}^{coc}, \alpha_{m,0}^{drink(01)}, \alpha_{m,0}^{drink(10)}, \alpha_{m,0}^{drink(11)}, \alpha_{m,0}^{mar(01)}, \alpha_{m,0}^{mar(10)}, \alpha_{m,0}^{mar(11)}, \alpha_{m,0}^{coc(01)}, \alpha_{m,0}^{coc(10)}, \alpha_{m,0}^{coc(11)})$

The likelihood of a sequence of drug j consumption indicators is written as in Model A, and is the same as equation 5. The individual contribution to the entire likelihood function is written as in equation 6.

IV.EMPIRICAL RESULTS

A. Parameter Estimates

Columns 1, 2 and 3 of Table 2 report parameters estimated by Model A corresponding to the latent utility (equation 1) of alcohol, marijuana and cocaine respectively. Columns 4, 5, and 6 are estimated by Model B, while columns 7, 8 and 9 correspond to Model C.[28]

Estimates for equation (1) yield a positive and statistically significant estimate for γ_{kk} (where k is alcohol, marijuana or cocaine), the coefficient associated with state dependence.

The coefficients associated with the stepping-stone effects show an interesting pattern. The stepping-stone effects from softer to harder drugs are positive and statistically significant[29] (e.g. alcohol has a stepping-stone effect on both marijuana and cocaine, while marijuana has a stepping-stone effect on cocaine). On the other hand, it remains inconclusive as to whether harder drugs reinforce the use of softer drugs.[30] Models A, B and C present strong evidence that there is a positive stepping-stone effect that operates mainly from softer to harder drugs.

Model C presents second-order terms that are statistically significant. A second-order specification allows the first-order state dependence to differ depending on whether that drug was also consumed in the previous periods.

---

[28] I do not report the parameter estimates of the initial conditions equation (equation 2).
[29] I define the order from softest to hardest, with alcohol a softer drug than marijuana, and marijuana a softer drug than cocaine.
[30] Model A reports statistically insignificant reverse stepping-stone effects, while Model C reports negative reverse-stepping stone effects. Model B is the only model that estimates positive reverse stepping-stone effects that are positive but significantly smaller than the stepping-stone effects from softer to harder drugs. For instance, Model B estimates a stepping-stone parameter of 0.45 from alcohol to marijuana, while the stepping-stone effect from marijuana to alcohol is 0.18. While the magnitude of the reverse stepping-stone effects is unclear, there is strong evidence that the stepping-stone effects operate from softer to harder drugs.

Model B estimates positive structural correlation between drug-specific transitory shocks

$\left(cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{mar}\right) = 0.58, \; cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{coc}\right) = 0.38, \text{and } cov\left(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right) = 0.46\right).$ [31]

Finally, time-invariant preferences across drugs are positively correlated as well. That is,

individuals who have a high inherent propensity to use any particular drug also have a higher

inherent propensity to use other drugs, and this pattern holds across specifications. Figure 3

graphs the intercept of the latent utility for marijuana against the intercept of the latent utility for

alcohol and cocaine, where Panels A, B and C of Figure 3 correspond to Models A, B and C

respectively.[32]

While the structural estimates reveal the sign of state dependence and stepping-stone effects,

they are not directly useful in answering questions such as to what extent the probability of

consuming cocaine increases if the respondent consumes marijuana in the previous period,

holding preferences for all drugs constant. The marginal effects are more useful at answering

these questions.

The marginal effects[33] of the state dependence parameter reveal the change in the probability

P_{ijt} in response to an infinitesimal change in $Y_{i,j,t-1}$. Because $Y_{i,j,t-1}$ is a discrete variable, the

---

[31] I cannot interpret the positive correlation between the time-varying components across drugs as evidence that the drugs are complements, because that would require more frequent data than a yearly panel. While this positive correlation may absorb some of the true complementarity across drugs, I interpret it as merely a correlation between the transitory shocks across drugs. The probit model estimates parameters $\phi_{12} = 1.311$ $(SE = 0.077)$, $\phi_{13} = 0.803$ $(SE = 0.103)$ and $\phi_{23} = 0.989$ $(SE = 0.072)$. The correlations $cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{mar}\right)$, $cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{coc}\right)$, and $cov\left(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right)$ are by construction between -1 and 1.

$$cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{mar}\right) = -1 + 2 * \frac{\exp(\phi_{12})}{1 + \exp(\phi_{12})}$$
$$cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{coc}\right) = -1 + 2 * \frac{\exp(\phi_{13})}{1 + \exp(\phi_{13})}$$
$$cov\left(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right) = -1 + 2 * \frac{\exp(\phi_{23})}{1 + \exp(\phi_{23})}$$

[32] Figure 3 graphs the intercept for the latent utilities corresponding to equation (1). I do not show the graph with the intercepts for the latent utilities of the initial conditions equations; however, they are also positively correlated.

[33] For models A and C (Logit), I compute the marginal effects $\frac{\partial P_{ijt}}{\partial z_j}$ for each individual i and period t, where

$\frac{\partial P_{ijt}}{\partial z_j} = P_{ijt} * \left(1 - P_{ijt}\right) * \beta_j$. I compute the average from year 1999 to 2007 for Model A. For Model C, I compute

average marginal effects (averaged over people and over time) reported in Table 3 should be interpreted as only an approximation. For instance, the logit model estimates that the probability of consuming alcohol, marijuana and cocaine in time t increases by 20, 15 and 7 percentage points as a result of consuming alcohol, marijuana and cocaine in the previous period, respectively. The marginal effects of the stepping stone effects are significantly smaller than the state dependence effects. A more appropriate way to quantify the role of state dependence and stepping-stone effects is described in the next subsection.

B. Quantifying the Effects of True State Dependence and Stepping-Stone Effects

A key feature of my dynamic discrete choice models is that I can use them to estimate the fraction of serial persistence in drug use that is attributable to state dependence. Similarly, I can use the model to quantify the role of the stepping-stone effects in explaining why the probability of consuming harder drugs is higher among those who consumed softer drugs in the previous period. For example, the probability of consuming cocaine is higher among those who consumed alcohol or marijuana in the previous period. Similarly, the probability of consuming marijuana is higher among those who consumed alcohol in the previous period.

To quantify the role of state dependence, I simulate the difference between the probability of consuming a particular drug among those who also consumed it in the previous period and those who did not,

---

$\frac{\partial P_{ijt}}{\partial z_j}$ for years 2000 to 2007 and take the average. Because Model C includes an interacted term of first and second-order state dependence, I compute the marginal effect of first-order state dependence separately for periods where the second-order lagged outcome was 0 and for periods where the second-order lagged outcome was 1. For Model B, I compute $\frac{\partial P_{ijt}}{\partial z_j} = \phi\left(V_{ijt}(\alpha_i)\right) * \beta_j = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}[V_{ijt}(\alpha_i)]^2\} * \beta_j$ for years 1999 to 2007, and take the average. The marginal effects are computed using simulated data because a random effect needs to be assigned to compute P_{ijt}. I repeat the simulation process of computing $\frac{\partial P_{ijt}}{\partial z_j}$ ten times and take the average.

$$P[Y_{ijt} = 1|Y_{i,j,t-1} = 1] - P[Y_{ijt} = 1|Y_{i,j,t-1} = 0] \quad [34], \tag{14}$$

under the assumption that state dependence is non-existent. In the absence of alcohol state dependence, the measure of alcohol persistence drops from 54.55 to 23.57 percentage points, using the parameters estimated by Model A. Similarly, the measure of marijuana persistence drops from 57 to 24 percentage points, in the absence of marijuana state dependence. Finally, the measure of cocaine persistence drops from 38.84 to 10.81 percentage points, when cocaine state dependence is non-existent. These simulations suggest that state dependence explains more than half of the observed persistence in drug use, and this pattern is highly robust across specifications (Table 4, Panel A).

To quantify the role of the stepping-stone effect from marijuana to cocaine, I simulate the difference between the probability of consuming cocaine between those who consumed marijuana in the previous period and those who did not,

$$P[Y_{it}^{coc} = 1|Y_{i,t-1}^{mar} = 1] - P[Y_{it}^{coc} = 1|Y_{i,t-1}^{mar} = 0] \tag{15}$$

under the assumption that the relevant stepping-stone effect is non-existent. In the absence of stepping-stone effects from marijuana to cocaine, this difference decreases from 15.4 to 11.37 percentage points, using the parameters estimated by Model A.

Similarly, turning "off" the stepping-stone effect from alcohol to cocaine decreases

$$P[Y_{it}^{coc} = 1|Y_{i,t-1}^{alc} = 1] - P[Y_{it}^{coc} = 1|Y_{i,t-1}^{alc} = 0] \tag{16}$$

---

[34] This difference is estimated separately for each year from 1999 to 2007. Next, I take the average over the nine periods after the initial period. If lagged consumption of drug j was randomly assigned, this difference would be interpreted as the causal effect of lagged consumption of drug j on current consumption of drug j. However, lagged drug use is not randomly assigned, and is highly driven by individual preferences. Let us focus on equation 14. The difference in probabilities provided by equation 14 is driven partially by individual preferences and partially by state dependence for drug j. I estimate the gap in equation 14 using real data. Then, I use the parameter estimates to see whether the model simulates the actual data closely. Finally, I also simulate this measure of persistence after I turn off the state dependence parameter ($\gamma_{jj} = 0$), while holding all other estimated parameters constant. This simulation estimates a counterfactual measure of drug j persistence in the absence of state dependence of drug j. A limitation to this approach to quantifying the effects of state dependence within and between drugs is that it is an unlikely situation that the relevant parameter γ_{kj} would change while all other parameters remain constant.

from 6.78 to 4.66 percentage points, using parameters estimated by Model A.

Finally, after turning "off" the stepping-stone effects from alcohol to marijuana, simulating the difference between the probability of consuming marijuana among those who consumed alcohol in the previous period and those who did not decreases the following difference,

$$P[Y_{it}^{mar} = 1|Y_{i,t-1}^{alc} = 1] - P[Y_{it}^{mar} = 1|Y_{i,t-1}^{alc} = 0], \tag{17}$$

from 21.87 to 17.89 percentage points, using the parameters estimated by Model A. These estimates are robust across specifications, suggesting that stepping-stone effects have only a modest role in explaining why individuals who use soft drugs are more likely to consume harder drugs in the next period (Table 4, Panel B).

C. Evaluating the Model

This section evaluates whether the model predicts the distribution of drug histories and the distribution of contemporaneous consumption bundles.

Table 5 compares the predicted and actual share of the sample that belongs to each of the mutually exclusive cells, defined by the number of periods in which each drug was used and the number of transitions from use to non-use and vice-versa.[35] I collapse the 2^{10} possible drug histories that can arise for each drug j, some of which have a negligible sample size, into 22 cells for alcohol and marijuana, and 10 cells for cocaine.

Columns 3 and 4 report the correspondence between actual and predicted histories for alcohol (Panel A), marijuana (Panel B), and cocaine (Panel C) by Model A. Columns 5 and 6 report such a correspondence for Model B, while columns 7 and 8 report this correspondence for Model C.

---

[35] For instance, an individual whose alcohol sequence for the entire 10 periods is (0001110000) belongs to the cell with 3 periods of use and 2 transitions.

The estimated models reflect the following drug use patterns. First, the largest cell in alcohol use is composed of individuals who used alcohol the entire 10 periods (cell of 8-10 periods of drug use, and zero transitions) while the largest cell in marijuana and cocaine use is composed of individuals who have never used it (cell of zero periods of drug use).

To informally measure the goodness of fit, I construct the Pearson Goodness of Fit Statistic for each drug j

$$PCGF_j = \sum_{k=1}^{K_j} \frac{(O_{kj} - E_{kj})^2}{E_{kj}} \tag{18}$$

where O_{kj} is the number of observations in cell k for drug j, E_{kj} is the number of predicted observations in cell k for drug j, and K_j is the number of cells for drug j.[36] Table 2 reports the corresponding $PCGF_j$ for alcohol, marijuana, and cocaine. [37]

The probit specification allows for correlated transitory shocks across drugs, which is advantageous when predicting contemporaneous consumption bundles. Table 6 provides an assessment of how well these models predict contemporaneous drug consumption bundles by year. Columns 3 and 4 report the correspondence between actual and predicted consumption bundles[38] in each period using parameter estimates from Model A. Columns 5 and 6 correspond to Model B, and columns 7 and 8 correspond to Model C.

---

[36] The Pearson Goodness of Fit Statistic is an informal summary measure of fit, which is based on the deviation between actual and predicted observations per cell. Grouping participation sequences into mutually exclusive cells is a common approach when the researcher is faced with the challenge of predicting a very large number of sequences, where some of those sequences have almost negligible frequencies (Card and Hyslop, 2005; Hyslop, 1999; and others). For multinomial outcomes, the PCGF can have a chi-square distribution and hence it is a formal statistic as opposed to an informal diagnostic (Moore 1977; Andrews 1988)

[37] The PCGF is calculated with cells different than those reported in Table 5 due to the small cell sizes of some combinations of periods of use and transitions. Tables with the cells that were used to calculate the PCGF to assure large enough sizes in each cell can be provided upon request. Models A and C combined cells to guarantee they had at least 40 observations, while Model B combined cells to guarantee they had at least 30 observations. Model A and Model C used the same cells to compute the PCGF, which makes their PCGF estimates comparable.

[38] To compare the actual and predicted bundles by year, I divide each period's choices of alcohol, marijuana, and cocaine into 8 mutually exclusive cells: (0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,0), (0,1,1), (1,0,1), (1,1,1). The first

Each bundle represents a possible combination of three binary variables at each period,

$(Y_{i,t}^{drink}, Y_{i,t}^{mar}, Y_{i,t}^{coc})$, where Y_{ijt}[39] has a value of 1 if respondent i consumed drug j in time t and a

value of 0 otherwise.

These three models reflect the following patterns in drug consumption bundles. First, the

number of individuals consuming (0,0,0) decreases over time. Second, among bundles where

only one drug was consumed at time t, the most popular drug to consume was alcohol, followed

by marijuana, followed by cocaine. Third, among bundles where two drugs were consumed at

time t, the most popular combination was alcohol-marijuana, followed by alcohol-cocaine,

followed by marijuana-cocaine.

Simply by looking at table 6, one observes that, while the logit models predict

contemporaneous drug consumption bundles fairly well (Model A and C), the probit

specification (Model B) predicts contemporaneous consumption bundles just as well, with only

three mass points, and with only first-order state dependence.

D. SPECIFICATION DIAGNOSTICS

A wide range of residuals have been proposed for non-linear models (Gourieroux et al

1987; McCall 1994; Chesher and Irish 1987). Given that the generalized residuals in nonlinear

models depend on unobservable latent variables, they are approximated by the difference

between the actual and the expected value, normalized for heteroskedasticity.[40]

---

component of the vector is an indicator for alcohol use, the second component corresponds to marijuana, and the third corresponds to cocaine.

[39] Y_{ijt}, ε_{ijt} are written as $Y_{it}^{drink}, Y_{it}^{mar}, Y_{it}^{coc}, and\ \varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}$ when j={drink, mar, coc}

[40] The generalized residuals are the difference between the actual binary variable Y_{ijt} and its expected value, normalized to correct for heteroskedasticity. Using the unconditional probabilities estimated by maximum likelihood (the share of the sample that belongs to each type without taking into account their observed sequence), I compute type-specific mean generalized residuals. Then, I estimate the generalized residuals using the posterior distributions of the random effects as weights for the generalized residual evaluated at each mass point. The posterior distributions of the random effects can be calculated using the Bayes rule, the observed outcome and the unconditional probabilities of being each type. Refer to Appendix A for a description of how these generalized

To examine the extent to which the identification assumptions required by Model A cause misspecification, I present estimated means, variances, and 1st-5th order autocorrelations of generalized residuals, as well as correlations across drugs. Comparing whether the sample-analogues of conditions 1 to 5 are close to their expected value under the null hypothesis of a correctly specified model allows me to choose a model with relatively little serial correlation and cross-drug contemporaneous correlation in the predicted errors.[41]

$$E\left[r_{i,t}^{j}(\alpha_i)\right] = 0 \tag{19}$$

$$E\left[r_{i,t}^{j}(\alpha_i)^2\right] = 1 \tag{20}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5} \tag{21}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t}^{j'}(\alpha_i)\right] = 0 \tag{22}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j'}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5} \tag{23}$$

Table 7 shows that the mean $E\left[r_{i,t}^{j}(\alpha_i)\right]$ and variance $E\left[r_{i,t}^{j}(\alpha_i)^2\right]$ of the generalized residuals are zero and one respectively, even in the simplest logit model with only first-order state dependence. The predicted errors have small but statistically significant serial correlation in the predicted errors $E\left[r_{it}^{j}(\alpha_i), r_{i,t-k}^{j}(\alpha_i)\right]$ for k=1,2,3,4,5. For instance, Model A estimates $E\left[r_{it}^{coc}(\alpha_i), r_{i,t-1}^{coc}(\alpha_i)\right] = -0.0089$ with SE 0.0034.[42] A model with second-order state

---

residuals were constructed, and refer to Gourieroux et al (1987) for a thorough description of generalized residuals for non-linear models.

[41] Card and Hyslop (2005) implemented the first three generalized residual sample-analogue diagnostics to diagnose serial correlation in the predicted errors. These three equations were enough to diagnose misspecification because they evaluated persistence of a single binary variable (to participate or not to participate in welfare). My study includes three binary variables (to use or not to use alcohol, marijuana, and cocaine), which requires extending these diagnostics to evaluate the last two sample-analogue generalized residuals.

[42] I report the $SE = \frac{SD}{\sqrt{N}}$ in Table 7, where SD is the standard deviation of the original random variables $r_{it}^{j}(\alpha_i)$, $r_{it}^{j}(\alpha_i)^2$ and $[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j}]$ respectively and SE is the standard error of the following sample average (the generalized-residuals diagnostics) : $E[r_{it}^{j}(\alpha_i)]$, $E[r_{it}^{j}(\alpha_i)^2]$, and $E[r_{it}^{j}(\alpha_i), r_{i,t-k}^{j}(\alpha_i)]$, respectively.

dependence (Model C) reduces the extent to which first and second-order serial correlation for all drugs are a source of misspecification.

Panel B of Table 7 shows evidence of correlated predicted errors "between" drugs, as $E[r_{it}^{drink}(\alpha_i), r_{i,t}^{mar}(\alpha_i)]$, $E[r_{it}^{drink}(\alpha_i), r_{i,t}^{coc}(\alpha_i)]$, and $E[r_{it}^{mar}(\alpha_i), r_{i,t}^{coc}(\alpha_i)]$ are significantly different from zero. This is clear evidence of misspecification for the logistic models, but not for the probit specification (Model B). [43]

Finally, I show the cross-time, cross-drug correlation sample-analogue generalized residual $E[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j'}(\alpha_i)]$ where k=1,2,3,4,5 and j,j'={alcohol, marijuana, cocaine}.Table7 Panel B shows evidence that the assumption of uncorrelated cross-time cross-drug errors does not pose a threat of misspecification and hence it is not necessary to relax that assumption by incorporating higher order stepping-stone effects.

V. HETEROGENEOUS STATE DEPENDENCE AND STEPPING-STONE EFFECTS

To improve upon the previous specifications, which assume homogeneous state dependence and stepping-stone effects, I estimate trivariate logit models that permit the degree of state dependence and stepping-stone effects to vary by age (Model D), by gender (Model E), by inherent propensity to participate in drug use (Model F and G), by intensity of drug use (Model H), and by accumulation of addiction capital (Model I).

A. Model D: Do Within and Between State Dependence Vary with Age?

---

[43] The correlation of the generalized residuals $E[r_{it}^{drink}(\alpha_i), r_{it}^{mar}(\alpha_i)] = 0.1759$, and $E[r_{it}^{drink}(\alpha_i), r_{it}^{coc}(\alpha_i)] = 0.0744$, and $E[r_{it}^{mar}(\alpha_i), r_{it}^{coc}(\alpha_i)] = 0.173$, reported in Panel B on Table 7, do not match the correlation of the structural residuals reported in Table 2 correspondingly, where $p_{12} = cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}) = 0.5754$, $p_{13} = cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{coc}) = 0.3812$, and $p_{23} = cov(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}) = 0.4578$. Simulated data shows that the correlation of generalized residuals is always lower than the correlation of structural residuals. While the computed correlation of generalized residuals is not relevant on its own, it is zero when the correlation of structural residuals is zero, and non-zero when the correlation of structural residuals is non-zero. Panel B of Table 2 indicates that the correlation of structural residuals is positive and statistically significant, which is in line with the correlation of generalized residuals reported in Table 7.

I consider a more general model that permits the state dependence and stepping-stone effects to vary by the age of the respondent. This specification relaxes the assumption of "linear in log odds" of Model A by replacing γ_{kj} in equation (1) with

$$\gamma_{kj=}(\gamma_{kj}^{cons} + \gamma_{kj}^{age} age_{it}) \tag{24}$$

Table 8 reports that state dependence increases with age for all three drugs, while stepping-stone effects decrease with age.[44] In particular, the state dependence parameters for each drug are as follows:

$\gamma_{drink,drink=}$ (-0.50+0.12*age_{it})

$\gamma_{mar,mar=}$ (-0.32+0.11*age_{it})

$\gamma_{coc,coc=}$ (-0.25+0.10*age_{it})

On the other hand, the stepping-stone effects from softer to harder drugs decrease with age

$\gamma_{drink,mar=}$ (2.23-0.11*age_{it})

$\gamma_{drink,coc=}$ (1.79-0.09*age_{it})

$\gamma_{mar,coc=}$ (1.32-0.04*age_{it})

as do the "reverse" stepping stone-effects

$\gamma_{mar,drink=}$ (0.93-0.05*age_{it})

$\gamma_{coc,drink=}$ (1.79-0.10*age_{it})

$\gamma_{coc,mar=}$ (1.02-0.04*age_{it})

This model presents strong evidence that the habit of using a particular drug may be harder to break with age. On the other hand, the stepping-stone effect decreases with age, indicating that soft drugs are more likely to have a stepping-stone effect to harder drugs at early ages.

---

[44] This is reflected by a positive γ_{kk}^{age} and a negative γ_{kj}^{age} when k≠j.

While previous literature has evaluated the effect of consuming alcohol and marijuana at early ages on educational attainment (e.g., Van Ours and Williams 2009; Register et al. 2001; Yamada et al. 1996), only few studies analyze their effect on future use of hard drugs. After separating environmental factors from the true effect of early marijuana consumption on cocaine, Lynskey et al. (2006) claim that individuals who consumed marijuana by age 18 are more likely to consume cocaine in the future than their same-sex twin who either consumed marijuana after age 18 or did not consume it at all. Furthermore, Yu and Williford (1992) claim that consuming alcohol at early ages, particularly between ages 13 and 15, increases the probability of progression to marijuana. My study is in line with the previous literature and points out that the stepping-stone effects are higher at younger ages

B. Models E: Do Within and Between State Dependence Vary by Gender?

First, I explore an alternative model where I include interaction terms within and between) state dependence and gender,

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{male} male) \tag{25}$$

Columns 4,5 and 6 of Table 8 (Model E) report γ_{kj}^{cons} and γ_{kj}^{male}. While gender does not define a clear pattern, the state dependence parameter for alcohol and marijuana is higher for men than for women. On the other hand, there are no gender-differences in cocaine state dependence. Men have a higher stepping-stone effect from alcohol to cocaine, while women have a higher stepping-stone effect from cocaine to alcohol.

C. Model F: Do Within and Between State Dependence Vary by α_j?

In this section, I allow the stepping-stone effect from drug k to j (γ_{kj}) to vary with the random effect associated with drug j (α_j).

This specification enables me to evaluate whether individuals with a high taste for a particular drug are more or less likely to be influenced by past consumption of other drugs. For instance, past marijuana use may not have much of an effect on future cocaine use among those with high preferences for cocaine, since they were going to consume cocaine regardless. Allowing the model to capture such behavior requires the following specification

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{\alpha_j}\alpha_j) \tag{26}$$

Columns 7, 8 and 9 of Table 8 report that the state dependence for alcohol is lower for those with higher time-invariant preference for alcohol, while cocaine state dependence is higher among those with high inherent propensity to consume cocaine. State dependence for marijuana is unaffected by time-invariant preferences for marijuana. Regarding the stepping-stone effects, respondents with high preference for cocaine (α_{coc}) are more easily influenced to consume cocaine by past alcohol consumption (higher $\gamma_{drink,coc}$) but less influenced by past marijuana consumption (lower $\gamma_{mar,coc}$) than their counterparts with lower α_{coc}. Respondents with high preferences for marijuana (α_{mar}) are more easily influenced to consume marijuana by past consumption of alcohol (higher $\gamma_{drink,mar}$) than respondents with low α_{mar}. Finally, respondents with high preferences for alcohol (α_{drink}) are more easily influenced to drink alcohol by past marijuana and cocaine use than those with low α_{drink}.

D. MODEL G: Do Within and Between State Dependence Vary by α_k?

Finally, I explore a model where I include interaction terms between the stepping-stone effects from drug k to j (γ_{kj}) and the random effects associated with drug k (α_k). Allowing the within and between state dependence to vary with α_k enables me to capture a plausible scenario where, for instance, consuming marijuana in the current period might make a

respondent with high inherent preference for marijuana less curious about consuming any other drug in the future.

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{\alpha_k} \alpha_k) \tag{27}$$

Columns 10, 11 and 12 of Table 8 report the following patterns: (1) Respondents with higher preferences for alcohol (α_{drink}) have higher stepping-stone effects from alcohol to marijuana and from alcohol to cocaine, but a lower alcohol state dependence than their counterparts with lower α_{drink}. (2) Respondents with high preferences for marijuana have higher stepping-stone effects from marijuana to alcohol and lower stepping-stone effects from marijuana to cocaine than their counterparts with lower α_{mar}. (3) Finally, individuals with high preference for cocaine α_{coc} have a higher stepping-stone effect from cocaine to alcohol and a higher cocaine state dependence than their counterparts who have lower α_{coc}.

E. Model H: Does Within and Between State Vary with Intensity of Use?

The preceding framework allowed for only one level of drug use, without distinguishing heavy users from non-heavy users. Orphanides and Zervos (2003) highlight the role of individual learning about drug-specific addiction levels and the individual's own "addictive tendencies" on the individual's resulting experimentation with drugs. According to their model, individuals who learn their "addictive tendencies" after a critical point will become drug addicts, while those who learn it before will avoid becoming addicted.

In order to allow for experimentation with drugs to influence more intense use in the future, I allow for state dependence to vary with intensity of use. Furthermore, I analyze the extent to which experimentation or frequent use of a particular drug influences the use of other drugs – in other words, evaluating how the stepping-stone effects vary by intensity of use.

To improve upon the previous models, I estimate an ordered logit model with three potential outcomes for each drug in order of intensity by $\{0,1,2\}$, which correspond to not using drug j, using low levels of drug j, and using high levels of drug j in the last year.[45]

The latent utilities for each drug now differentiate lagged low levels of use from lagged high levels of use.[46] The latent utilities for each drug j are as follows

$$U_{ijt} = \underbrace{\delta_{1j}(t - t_0) + X_{it}\beta_j + \sum_{k=1}^{J} \gamma_{kj}^{Low} Y_{i,k,t-1}^{Low} + \sum_{k=1}^{J} \gamma_{kj}^{High} Y_{i,k,t-1}^{High} + \alpha_{ij}}_{V_{ijt(\alpha_{ij})}} + \varepsilon_{ijt} \qquad (28)$$

$$U_{ij0} = X_{i0}\lambda_{0j} + \alpha_{ij0} + \varepsilon_{ijt}^{initial} \qquad (29)$$

I assume that the utility of not using drug j in period t is zero. At each period, Y_{ijt} has a value of 0, 1 or 2 if the following equations hold, respectively

$$Y_{ijt} = \begin{cases} 0, & U_{ijt} \leq c_1 \\ 1, & c_1 \leq U_{ijt} \leq c_2 \\ 2, & U_{ijt} \geq c_2 \end{cases} \qquad (30)$$

This model estimates the thresholds c_1 and c_2, along with the parameters of the latent utilities $V_{ijt}(\alpha_j)$. With a logistic error assumption, the probability of an individual using drug j in low or high levels or abstaining from using drug j has a closed form solution and can be written as follows:

$$P\left(Y_{ijt} = 0 \middle| \alpha_{ij}\right) = \frac{\exp\left[c_1 - V(\alpha_{ij})\right]}{1 + \exp\left[c_1 - V(\alpha_{ij})\right]} \qquad (31)$$

$$P\left(Y_{ijt} = 1 \middle| \alpha_{ij}\right) = \frac{\exp\left[c_2 - V(\alpha_{ij})\right]}{1 + \exp\left[c_2 - V(\alpha_{ij})\right]} - \frac{\exp\left[c_1 - V(\alpha_{ij})\right]}{1 + \exp\left[c_1 - V(\alpha_{ij})\right]} \qquad (32)$$

---

[45]I define low levels of alcohol and marijuana use as consumption at least once in the last year but at most five days in the last month. Similarly, high levels of alcohol and marijuana use are defined as consumption more than five days in the last month. Because the NLSY97 does not include last month measures of cocaine use, I define low levels of cocaine use as consumption at most five times in the last year, and high levels of cocaine use as consumption more than five times in the last year.

[46] The latent utility for the ordered logit model does not include an intercept because the intercept would not be identified separately from the threshold. Also, the random effect affects the intercept, and allows the threshold to vary by type.

$$P\left(Y_{ijt} = 2 \middle| \alpha_j\right) = 1 - \frac{\exp\left[c_2 - V(\alpha_j)\right]}{1 + \exp\left[c_2 - V(\alpha_j)\right]} \tag{33}$$

The likelihood of a sequence of drug j consumption indicators is written as in Model A; it is the same as equation 5. The individual contribution to the entire likelihood function is written as in equation 6.[47]

The last three columns of Table 8 present the parameter estimates of equation (28), which specifies the latent utilities for alcohol, marijuana and cocaine, respectively. State dependence is positive and statistically significant for all three drugs, with a slightly larger effect when the drug was consumed in "high levels" in the previous period.

Similarly, the stepping-stone effects from softer to harder drugs are positive and statistically significant for all drugs, with a relatively larger effect when the source drug was consumed in "high levels" rather than "low levels." For instance, the stepping-stone effect from marijuana to cocaine is higher among respondents who used "high levels" of marijuana in the previous period.

Regarding the stepping-stone effect from harder to softer drugs, lagged marijuana consumption does not affect current alcohol consumption, regardless of the intensity of past marijuana use. Interestingly, a high level of past cocaine use increases future alcohol use with no effect on future marijuana use. On the other hand, low levels of past cocaine use increases future consumption of marijuana but not alcohol in the future period.

F. <u>MODEL I: Do Within and Between State Dependence Vary with Years of Accumulation?</u>

In this subsection, I allow state dependence (within and between) from drug k to j (γ_{kj}) to vary with the number of years that have elapsed since the first time the respondent used drug k (acc_{kt})[48]

---

[47] The only difference is that Y_{ijt} is no longer binary and now has three values: $\{0,1,2\}$.

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{acc_k} * acc_{kt}) \tag{34}$$

Previous literature studied the role of cumulative measures of past drug consumption on current consumption of such drug (Becker, Grossman, and Murphy 1991; Becker and Murphy 1988; Pollack 1970; Pollack 1976; Stigler and Becker 1977), while overlooking the effect of past drug accumulation of a particular drug on current consumption of other drugs.

Accumulation of addiction capital influences current substance consumption through two channels. On one hand, the longer an individual has consumed a particular drug, the more pleasure he or she derives from current consumption[49]. On the other hand, greater past consumption can lower current utility for a given amount for addictive goods[50].

The state dependence (within and between) parameters reported in the last three columns of Table 8 show that state dependence increases with accumulation of addiction capital ($\gamma_{kj}^{acc_k}>0$ for k=j and j,k={alcohol, marijuana, cocaine}), and these estimates are statistically significant. While accumulation of capital defines a clear pattern for within-drug state dependence, it does not for between-drug state dependence. The effect of accumulation of addiction capital on the stepping stone parameters is statistically significant at a conventional level only for the alcohol-marijuana pair. The stepping stone effect from alcohol to marijuana decreases with the accumulation of alcohol capital and the stepping stone effect from marijuana to alcohol decreases with accumulation of marijuana capital.

G. What about Drug and Year Specific Prices and Cost of Accessibility?

---

[48] Accumulation (or stock) of addiction capital acc_{kt} measures years of exposure to drug k and is computed as the number of years that have elapsed between year t and the age of initiation to drug k.

[49] This can be driven by a variety of reasons such as habit formation or simply learning how to use a particular drug. For instance, individuals are more likely to cough the first times they use cigarettes than after being smokers for several months.

[50] Becker, Grossman and Murphy (1991) refer to these effects as reinforcement and tolerance. In their study, accumulation of addiction capital affects current utility for a given level (tolerance) and also current consumption level (reinforcement).

A limitation of the previous specifications is the lack of exclusion restrictions, that is a drug and year specific variable (Z_{ijt}) that directly enters the utility of drug j at time t (U_{ijt}) but does not enter the utility for any other drug j' or period t'. While time-varying drug prices seem to be the obvious answer, the full price of consuming drugs consists of the monetary price and the non-monetary component associated with ease of access to the drug.[51]

Previous literature on substance use has documented the strong association between prices of drugs and their consumption level, especially among youth (Dhaval 2005; Saffer and Chaloupka 1999; Grossman and Chaloupka 1998; Chaloupka, Grossman and Tauras 1999, among others) and relies on this association to study the effects of hard drugs on risky and criminal behavior (Brown and Silverman 1974; Silverman and Spruill 1977; Dobkin and Nicosia 2009; Markowitz 2000, among others).

To measure the price of alcohol, I use the series of price of malt beverages per 16 oz. provided by the Bureau of Labor Statistics by region and month. I compute a yearly average and convert it to1980-1984 dollars.[52]

I obtain year specific prices per expected pure gram for a variety of hard drugs such as powder cocaine, crack cocaine, heroin, methamphetamine, and price per bulk gram of marijuana from the 2008 ONDCP report entitled "The Price and Purity of Illicit Drugs: 1981-2007"[53]

---

[51] While price is a natural choice for an exclusion restriction, any time-varying drug specific variable fits the criteria to be an exclusion restriction. Gentzkow (2007) uses non-price variables (indicators for whether the individual has internet access at work, or uses the internet for work related or education related tasks, among other non-price variables) as exclusion restrictions for the utility of reading online papers, given the lack of variation in the price of reading newspapers online (price is zero and it does not vary over time).

[52] While most of the variation in prices of alcohol across states is driven by alcohol taxes, state level beer taxes do not exhibit much variation over time. The federal excise tax on beer has not changed since 1991. There is very limited within state variation in the beer tax level, which makes beer tax a problematic exclusion restriction (Carpenter et al 2007). Price is a more attractive choice for alcohol exclusion restriction.

[53] Office of National Drug Control Policy, 2008, The Price and Purity of Illicit Drugs: 1981-2007. Institute for Defense Analysis (IDA) Paper P-4369, electronically accessible through the following World Wide Web address http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/bullet_1.pdf . The accompanying *Technical Report* is available at http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/price_purity_tech_rpt07.pdf .

prepared by the Institute for Defense Analysis (DEA) for the Office of National Drug Control Policy (ONDCP). Estimates of price and purity presented in their report were obtained from records in the System to Retrieve Information from Drug Evidence (STRIDE) database.[54] The prices reported in STRIDE are collected by undercover DEA, FBI and state and local police narcotics during undercover seizures in 139 cities and reflect actual street prices because inaccurate price offers would make drug dealers suspicious, threatening the safety of undercover agents (Taubman, 1991).

A problem of the STRIDE data is that the total cost of purchasing a particular amount is not proportional to weight. Price per gram and level of purity varies by purchased amount. To overcome this problem, the 2008 ONDCP reports national quarterly EPH (expected purity hypothesis[55]) estimates for each illicit drug for three or four quantity levels[56]. Due to the lack of purity data for marijuana, I only have access to price per bulk gram of marijuana by purchased amount (0.1 to 10 grams, 10-100 grams, more than 100 grams).[57]

The full cost of using drugs is not only reflected by prices but also by ease of drug accessibility. The variable pop_{ijt} provides a proxy for drug prevalence, and measures the state level deviations from national trends in treatment admissions for substance abuse for a particular

---

[54] STRIDE is a dataset maintained by the DEA and provided to ONDCP and IDA

[55] Office of National Drug Control Policy (2008) define the "Expected Purity Hypothesis" model as "The output of the EPH modeling can be viewed as regression-based estimates of the expected purity and expected price per pure gram for distinct combinations of illicit drugs, quantity levels, and geographical locations. National indices are constructed as population-based weighted averages of results across different cities and census regions. In the estimation of prices, each STRIDE transaction price is normalized by the local value of the expected purity, vice by the assayed purity of the specific transaction."

[56] For instance, the reported price per expected pure gram of crack cocaine is reported separately for purchases between 0.1 to 1 gram, 1 to 15 grams, and more than 15 grams.

[57] An alternative dataset for marijuana prices is provided by the *High Times* magazine, which compiles prices voluntarily posted by individuals who recently purchased marijuana. While marijuana seizures only constitute 3% of the STRIDE database, these are enough observations to measure price variation over time (I do not need marijuana prices by state and year). The STRIDE marijuana prices must reflect market prices in order to not look suspicious during busts and seizures.

drug, for the state in which respondent i resided in year t[58]. The numerator measures the share of admissions to a substance abuse treatment facility who reported abusing drug j in state s and year t (state s is the state of residence for individual i in year t), as indicated in the TEDS-A dataset. The denominator is the national share of admissions to a substance abuse treatment facility in year t that reported abuse of drug j.

$$pop_{ijt} = \frac{\# ppl\ TEDS_{j,s,t} / \# ppl\ TEDS_{s,t}}{\# ppl\ TEDS_{j,USA,t} / \# ppl\ TEDS_{USA,t}} \tag{35}$$

Respondents with $pop_{i,j,t}$ greater than 1 live in a state that has a higher share of individuals admitted to a substance abuse treatment center for abusing drug j than the national share, and I interpret this as individual i having a lower cost of accessing drug j than average[59].

The model estimated in this section incorporates year and drug specific variables tv_{ijt} to the latent utilities (equation 1), as well as tv_{ij0} to the initial conditions equation (equation 2),

$$U_{ijt} = \delta_{0j} + \delta_{1j}(t - t_0) + X_{it}\beta_j + \sum_{k=1}^{J} \gamma_{kj} Y_{i,k,t-1} + tv_{i,j,t} + \alpha_{ij} + \varepsilon_{ijt} \tag{36}$$

$$U_{ij0} = \omega_{0j} + X_{i0}\lambda_{0j} + tv_{i,j,0} + \alpha_{ij0} + \varepsilon_{ijt}^{initial} \tag{37}$$

where $tv_{i,drink,t} = \theta_1 * I[Age_{it}\theta_1 \geq 21] + \theta_2 * Z_{i,drink,t}$, $tv_{i,mar,t} = \theta_3 * Z_{i,mar,t}$, $tv_{i,coc,t} = \theta_4 * Z_{i,coc,t}$, and Z_{ijt} is the drug and time specific exclusion restriction. For the initial conditions

---

[58] There are other potential measures of drug prevalence such as the number of drug-related emergency room visits and the share of employers who request drug testing prior to employment, but the years for which those datasets were available did not overlap with my sample.

[59] Note that $pop_{ijt}>1$ can also be interpreted as the individual living in a state that is very intolerant of drug j where users of drug j are more easily sent to substance abuse programs (in which case individual I faces a higher cost of accessing drug j). The sign of pop_{ijt} is irrelevant for this study. The variable pop_{ijt} enters the utility of alcohol and cocaine positively, while it enters the utility for marijuana negatively. Regardless of how I interpret pop_{ijt}, STRIDE prices are negatively correlated to the share of TEDS treatment center admissions, at least for cocaine (ONDCP IDA Paper P-3219), and hence they provide additional information on drug and year specific prices.

equations, I use $tv_{i,j,0} = \theta_{0j} * Z_{i,j,0}$ because every individual is under age 21 in the initial period. I explore with three specifications for $Z_{i,j,t}$.

The first three columns of Table 9 report state dependence (within and between) parameters (γ_{kj}) when $Z_{it} = pop_{ijt}$. State dependence within drug and stepping-stone effects from softer to harder drugs are positive and statistically significant. On the other hand, reverse stepping-stone effects from harder to softer drugs are statistically insignificant. These results remain robust to alternative specifications of Z_{it}.

Columns 4, 5, and 6 define $Z_{it} = (1/pop_{ijt}) * price_{ijt}$ where price for alcohol is the price of 16 oz of a malt beverage in the region individual i resided in year t in 1980-84 dollars. I use the price per gram of marijuana among purchases that were made in small volumes (0.1-10 grams) and the price of crack cocaine among purchases that were made in small volumes (0.1-1 gram). Finally, the last three columns use the same definition of $Z_{it} = (1/pop_{ijt}) * price_{ijt}$ but use the price per gram of marijuana among purchases that were made in large volumes (more than 100 grams) and the price per expected pure gram of crack cocaine that was purchased in large amounts (more than 15 grams). The results remain robust to my choice of price per gram of marijuana and cocaine[60].

Finally, I also included an indicator for whether the respondent is older than 21 years of age at the time of interview in $tv_{i,drink,t}$, and this positively affected the latent utility for alcohol.

H. Complements or Substitutes?

---

[60] I also tested models that include all possible combination of drug prices (price per gram of marijuana for purchases in low volumes-between 0.1-10 grams- and high volumes-more than 100 grams) and for hard drugs (powder cocaine, crack cocaine, heroin, and methamphetamine when purchased in small and large volumes). The results remained robust to these 16 specifications.

An analogous problem to separating state dependence from unobserved heterogeneity, it

is to separate complementarity in utility (i.e. the utility from consuming two different drugs is

higher than the sum of the utilities of consuming each of those drugs separately) from correlation

in consumer preference (individuals who have a high preference for any substance also have a

high permanent preference for other substances).[61]

Gentzkow (2007) develops a model to separate these two competing factors to explain

the correlation between contemporaneous purchases of print newspaper and consumption of

online newspapers. My paper also contributes to the literature by extending the model developed

by Getzkow (2007) to also allow for state dependence and stepping stone effects as well as

complementarity in utilities[62].

Previously, I estimated a multivariate probit model (Model B) where I allowed the

permanent as well as the transitory unobserved components of the latent utilities to be correlated

between drugs. A problem with the multivariate probit specification is that the utility from

consuming two drugs together is assumed not to have any value in addition to consuming both

drugs separately ($\Gamma_{ab} = 0$). This section relaxes that assumption.

Ideally, I would estimate a model where I define the utility for consuming drug j by itself

$U_{ijt} = V_{ijt} + \varepsilon_{ijt}$, then define the utility for consuming two drugs j and j' as $U_{ijj't} = V_{ijt} +$

---

[61] Table 6 shows that individuals are more likely to consume drugs in bundles, and this occurs particularly with harder drugs. Only 0.06% of my NLSY97 sample in 2007 consumed cocaine by itself, only and 0.04% consumed cocaine and marijuana alone, while 3% consumed alcohol, marijuana and cocaine. Similarly, 0.7% consumed marijuana alone while 14% consumed alcohol and marijuana. On the other hand, 59% consumed alcohol alone (Table 6).

[62] Gentzkow (2007) models the utilities of consuming two non-mutually exclusive goods a and b as follows (U_0, U_a, U_b, U_{ab} correspond to the utilities for not consuming either a or b, consuming only a, only b, and consuming both a and b, respectively). The parameter Γ determines whether a and b are complements ($\Gamma > 0$) or substitutes ($\Gamma < 0$) in utility.

$$U_0 = 0$$
$$U_a = \delta_a - \alpha p_a + v_a$$
$$U_b = \delta_b - \alpha p_b + v_b$$
$$U_{ab} = U_a + U_b + \Gamma$$

$V_{ij't} + \Gamma^{jj'} + \mathcal{E}_{ijj't}$ and the utility from consuming all three drugs j, j', and j'' as $U_{ijj'j''t} =$

$V_{ijt} + V_{ij't} + V_{ij''t} + \Gamma^{j,j'j''} + \varepsilon_{ijj'j''t}$ for all eight possible bundles of consumption of alcohol,

marijuana and cocaine. However, this model is impossible to estimate because there are some

bundles of consumption where the sample size is negligible.[63]

Because of that, I estimate these models separately in pairs, and establish whether they

are complements or substitutes. First, I estimate these models using the same $U_{ijt} = V_{ijt} + \mathcal{E}_{ijt}$

as defined in equation 1 and 2 (Panel A Table 10). Second, I use latent utilities that incorporate

drug and year specific exclusion restrictions as in equation 36 and 37 (Panel B Table 10). For

both specifications, I estimate a separate $\Gamma_0^{jj'}$ for each pair at the initial conditions.

This second specification with exclusion restrictions is equivalent to the model developed

by Gentzkow (2007) after having incorporated state dependence within and between drugs.

The state dependence and stepping stone parameters have the same interpretation as

before. Alcohol and marijuana are complements. While alcohol has a stepping stone effect on

marijuana, marijuana consumption has a negative stepping stone effect on alcohol.

The third and fourth columns report the relationship between alcohol and cocaine.

Alcohol has a positive and statistically significant stepping stone effect on cocaine, while lagged

cocaine use has no effect on current alcohol consumption.

Finally, the last two columns show the relationship between marijuana and cocaine.

While marijuana and cocaine are complements in utility, marijuana does not have a stepping

stone effect on cocaine. On the other hand, cocaine has a negative stepping stone effect on

---

[63] I estimate a similar model where the outcomes are (1) alcohol only, (2) marijuana only, (3) alcohol and marijuana, (4) cocaine alone, cocaine and alcohol, cocaine and marijuana, and (5) alcohol, marijuana and cocaine. While the state dependence and stepping stone parameters are not readily interpretable in that specification, that model estimates a positive and statistically significant $\Gamma^{jj'j''}$, which implies that there is added value to consuming all three drugs together. These estimates can be provided upon request.

marijuana. These results are robust to the inclusion of drug and time specific exclusion restrictions. The main purpose of this section was to separate complementarity in utility from correlation in preferences. Because I could not include lagged outcomes for all three drugs, I do not make these findings on the state dependence and stepping stone effects the preferred interpretation for such parameters.

I. POLICY CONSEQUENCES OF EARLY DRUG USE

In this section, I use the parameter estimates from the models to evaluate the effect of reducing the consumption of soft drugs on long term drug outcomes. Figure 4 compares the age profile of simulated cocaine consumption and four variants. These alternatives remove alcohol and marijuana consumption together until age 18, 20 and 22 respectively. I simply "turn off" consumption of alcohol and marijuana until the relevant ages. My simulations suggest that removing soft drugs until age 18 reduces long term cocaine use by large magnitudes, while removing cocaine for an additional two years, until age 20, has a milder impact. Finally, removing soft drugs until age 22 has almost negligible additional effects relative to removing them until age 20.[64] This is in line with a finding that stepping-stone effects decrease with age[65].

While removing alcohol and marijuana use

---

[64] Table A5 reports that the decrease in simulated cocaine consumption that occurs if alcohol and marijuana were artificially removed until age 18 is drastic and statistically significant until age 22. However, artificially removing alcohol and marijuana for two additional years until age 20 leads to a further statistically significant decrease in simulated cocaine consumption for the age group 21-22. Finally, artificially removing alcohol and marijuana for additional two years until age 22 leaves simulated cocaine consumption statistically indistinguishable from the previous counterfactual scenario.

[65] Table A5 presents standard errors for the estimates of the share of respondents who consume cocaine by age category that are presented in Figure 4. First, I simulate drug outcomes using the parameters estimated in the model and the demographics of my entire sample. Second, I use those simulated drug outcomes to obtain an estimate for the share of individuals in the sample that my model predicts will be using drugs by each age category (μ_{jatr} is the predicted share of individuals who consume drug j in age category a, and simulation repetition r for year t). Third, I repeat this process 100 times. Then, I compute a mean and standard deviation for $(\mu_{jat1}, \mu_{jat2}, \dots\dots, \mu_{jat100})$ and report those in Table A5.

entirely is not very realistic, it is more plausible to increase the cost of accessibility to encourage individuals who use the softer drugs to use them with moderation.[66] Figure 5 compares the share of respondents who consume cocaine and three variants. These alternatives assume that respondents who want to use alcohol or marijuana are somehow compelled to only use them in "low doses." Hence, I simply replace simulated high levels of consumption of alcohol and marijuana with low levels of consumption, instead of "turning them off."

Assuming that this change has no effect on the other parameters of the model, my structural estimates imply that regulating intensity of alcohol decreases simulated cocaine consumption (at both high and low levels) from 5.35% to 4.69%,[67] while regulating intensity of marijuana use decreases simulated cocaine consumption to 4.75%. These effects are comparable, but simulating policy changes to reduce intensity of alcohol consumption affects outcomes of more individuals than implementing policy changes to reduce intensity of marijuana consumption, because there are more heavy alcohol than heavy marijuana users.

Figure 6 illustrates the role of marijuana consumption in low amounts by comparing the simulated share of respondents who consume cocaine with two variants: one where simulated frequent users of marijuana are somehow compelled to consume only low levels of marijuana, and one where simulated frequent marijuana users are somehow entirely deterred from consuming marijuana. Simulations suggest that, for potential heavy marijuana users, eliminating marijuana use has a modestly greater effect on reducing cocaine use, relative to the effect of merely reducing levels of marijuana consumption. This gap decreases over time and becomes

---

[66] Carpenter and Dobkin (2009), for instance, presents evidence that a higher cost of accessing alcohol (provided by the minimum legal drinking age) reduces binge drinking and heavy use of alcohol.
[67] The share of individuals who use cocaine in a year (in low or high amounts) is 5.35%, averaged over years 1998 to 2007. When I simulate cocaine consumption (adding up low and high) under the counterfactual scenario where alcohol and marijuana in high amounts are replaced with low amounts, the share of cocaine users is 4.69%, averaged over years 1998 to 2007.

almost negligible as respondents become adults, which is in line with the finding that low

amounts of marijuana have a stepping-stone effect that operates mainly at young ages.[68]

These simulated scenarios highlight previously unappreciated benefits of policies aimed

at preventing alcohol and marijuana consumption. Because marijuana is already illegal, most

policies aimed at moderating consumption of soft drugs focus on alcohol regulation. I will

briefly discuss how the Zero Tolerance (ZT) Laws, the current minimum legal drinking age

(MLDA) of 21, and the MLDA increase in the 1980s are comparable to the simulated scenarios

described above.

Zero Tolerance Laws were implemented in all states by 1998, requiring suspension of the

driver's license of any driver under age 21 who has any amount of blood alcohol content; these

laws did not directly target drinking by individuals 21 and older. To the best of my knowledge,

there have been no studies directly evaluating the effect of ZT laws on marijuana and cocaine

consumption, but Carpenter (2004a) finds that they decreased binge drinking by 13% among

men younger than 21. My study suggests previously unappreciated benefits of policies that

prevent use of soft drugs at early ages, such as ZT laws.

---

[68] Table A6 presents standard errors for the estimates of the share of respondents who consume cocaine by year that were plotted in Figure 5 and 6. These estimates were computed in a similar way to the estimates presented in Table A5. I use the ordered logit model to compute these estimates, and the reported share of respondents who consume cocaine in year t corresponds to the sum of the share of respondents who consume cocaine in low levels and high levels. As we can see from Figure 5, there is a decrease in simulated cocaine users when we regulate intensity of both alcohol and marijuana consumption, and this decrease is statistically significant (Panel B versus Panel E of Table A6). However, the effects of artificially regulating either high levels of alcohol or high levels of marijuana consumption are statistically indistinguishable (Panel C and D of Table A6). I am not saying that the stepping stone effects from high levels of alcohol are the same as the effects of high levels of cocaine since there are more individuals using high levels of alcohol than there are using high levels of marijuana. Finally, figure 6 compares simulated cocaine consumption under the counterfactual scenario where high levels of marijuana consumption are replaced with low levels (Panel D of Table A6) versus a scenario where simulated frequent marijuana users are entirely deterred from consuming marijuana (Panel F of Table A6). While the second scenario leads to a lower share of cocaine users, they are statistically indistinguishable. Finally, Panel G reports simulated share of cocaine users under the assumption that all simulated marijuana users (frequent and infrequent) are deterred from consuming it, as a reference point.

A second policy that decreases frequency and intensity of alcohol consumption among youth is the current minimum legal drinking age of 21. Using the NLSY97, Yoruk and Yoruk (2011) found that all measures of alcohol consumption increased at age 21, causing an increase in the number of marijuana users but a statistically insignificant decrease in the frequency of use of marijuana. Because my model suggests that the stepping-stone parameters decrease with age, I would expect the increase in marijuana users associated with attaining the drinking age to be higher if the MLDA were lower.

A change in the MLDA occurred in 1983, when President Reagan imposed penalties on states that failed to increase the minimum legal drinking age from 18 to 21; all states had implemented the change by 1988. However, the effect of this policy on marijuana[69] is not comparable to my simulation exercise for two reasons. First, the increase in the MLDA took place in 1988 at the latest, and the NLSY97 interviews respondents were between 18 and 21 in 2003. Because factors that affect drinking have changed between 1988 and 2003, these two time periods are not comparable. Second, the increase in the MLDA decreased alcohol consumption only very slightly among respondents between 18 and 21 years old (DiNardo, Lemieux 2001). Because my simulation exercises artificially remove alcohol use entirely at early ages, or regulate intensity of alcohol use for everyone in my sample, these policies are not directly comparable.

Because of the illegal status of marijuana and cocaine, there are not many prevention policies targeted at particular age groups. Anderson et al. (2011) finds that legalization of medical marijuana increases marijuana consumption among adults but not minors. My study suggests

---

[69] DiNardo and Lemieux (2001) found that marijuana use increased slightly among respondents 18-21 who were affected by the increase in the minimum legal drinking age.

that the stepping-stone effects from medical marijuana to other drugs should not be very large in magnitude because adults have lower stepping-stone effects.

VI. CONCLUSION

This paper presents an explanation of two aspects of drug consumption. First, drug use exhibits persistence over time. That is, the probability of consuming a particular drug at any given period is higher among those who consumed that drug in the previous period. Second, the probability of consuming hard drugs is higher among those who consumed relatively softer drugs in the previous period.

In this paper, I develop a series of multiple-equation models that allow me to disentangle state dependence and stepping-stone effects from correlated unobserved heterogeneity. I find strong evidence of state dependence in drug use, as well as modest-sized stepping-stone effects that operate from softer to harder drugs. I also show evidence that both the permanent and transitory unobservable components of tastes are positively correlated across drugs. This positive correlation does not support an interpretation that drugs are complements to each other; such a finding would require more frequent data than the yearly NLSY97 survey.

A related finding is that the stepping-stone effects decrease with age, while state dependence increases with age, indicating that the habit of using a particular drug is harder to break for older individuals, and that consumption of softer drugs has a greater influence on consumption of harder drugs at early ages. I also consider an "ordered" model that allows me to distinguish between different levels of intensity of drug use at each point in time. An ordered logit model indicates that the stepping-stone effect from softer to harder drugs is stronger among those who consume high levels of the softer drug. This study suggests that one should take into

consideration age and intensity of soft drug consumption in investigating the effects of alcohol and marijuana use on long-run consumption of hard drugs.

This paper highlights the important role that individual preferences play in drug consumption. Most of drug persistence is explained by true state dependence within a drug type. On the other hand, only a small share of the gap between the conditional probabilities of consuming harder drugs among those who consumed softer drugs in the previous period versus those who abstained is explained by stepping-stone effects. While preventing alcohol and marijuana use has a modest impact on future cocaine use, these policies are constrained by the role of individual preferences.

While this study is not the ideal field experiment where drugs are randomized among individuals whose future drug consumption is followed by a researcher, taken together these specifications are much more informative than the raw data in assessing the degree to which illegal drugs exert a stepping stone effect.

APPENDIX A : Generalized Residuals Specification Diagnostics

In this section, I describe step-by-step how to build the generalized residuals that I use as specification diagnostics in section 4.

If the model is correctly specified, the following conditions must hold at the true location parameter of the time-invariant unobservable component (random effect), where r_{it}^{j} is a generalized residual

$$E\left[r_{i,t}^{j}(\alpha_i)\right] = 0$$

$$E\left[r_{i,t}^{j}(\alpha_i)^2\right] = 1$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t}^{j'}(\alpha_i)\right] = 0$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j'}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5}$$

A challenge to compute a sample-analogue of these five equations is that the random effect is unknown to the econometrician. Card and Hyslop (2005) implemented sample analogues to the first three equations to test the null hypothesis of the model being correctly specified. Because my study has three binomial variables, I need to extend their diagnostics to diagnose whether the last two equations hold under the assumption of the model being correctly specified.

Following Card and Hyslop (2005), I develop the generalized residuals using the posterior distributions of the random effects as weights for the generalized residual evaluated at each mass point.

Let π_k be the unconditional probability. Those probabilities are estimated by maximum likelihood and can be interpreted as the share of the sample that is type k, without taking into

account their observed sequence. Observing the outcomes, I can estimate the conditional or

posterior probabilities and the generalized residuals for the logit model as follows.[70]

Step 1: Assign the predicted probability of the observed sequence given a type

First, I write the likelihood for the observed outcome conditional on the estimated parameters $\hat{\beta}$

and conditional on the individual being type m. For logit, I write the likelihood as follows:

$$L_i(\alpha_m) = L_i^{drink}(\alpha_m) * L_i^{mar}(\alpha_m) * L_i^{coc}(\alpha_m)$$

where

$$L_i^j(\alpha_m) = P(Y_{ij0}|\alpha_m) \prod_{t=1999}^{2007} P(Y_{ijt}|\bar{Y}_{i,t-1}, \alpha_m)$$

$$\bar{Y}_{it} = (Y_{it}^{drink}, Y_{it}^{mar}, Y_{it}^{coc})$$

$$\alpha_m = (\alpha_m^{drink}, \alpha_m^{mar}, \alpha_m^{coc})$$

Step 2: Calculate the conditional or posterior probabilities of being each type

The conditional probabilities of being type m, given the observed outcome, can be calculated

using the Bayes rule, the observed outcome and the unconditional probabilities of being each

type.

$$w_i^m = P(\alpha_m|\bar{Y}_i) = \frac{L_i(\alpha_m)}{\sum_{m=1}^M \pi_m L_i(\alpha_m)}$$

Step 3: Build Generalized Residuals

$$r_{it}^j(\alpha_m) = \frac{Y_{ijt} - P_{ijt}(\alpha_m)}{\sqrt{[P_{ijt}(\alpha_m)][1 - P_{ijt}(\alpha_m)]}}$$

---

[70] The expected value is $E[Y_{ijt}]=P(Y_{ijt}=1)$ for the logit and probit specifications. I have also computed generalized residuals for the ordered logit model, where Y={0,1,2},

$E[Y_{ijt}]=0*P(Y_{ijt}^{none}) + 1 * P(Y_{ijt}^{low}) + 2 * P(Y_{ijt}^{high})$, and

$var(Y_{ijt}) = P(Y_{ijt}^{none}) * (0 - E[Y_{ijt}])^2 + P(Y_{ijt}^{low}) * (1 - E[Y_{ijt}])^2 + P(Y_{ijt}^{high}) * (2 - E[Y_{ijt}])^2$. I did not report in the paper the generalized residual diagnostics for models with heterogeneous state dependence but I can provide them upon request.

where Y_{ijt} is the observed outcome for drug j and $P_{ijt}(\alpha_m)$ is the predicted probability of using

drug j conditional on the estimated parameters, lagged drug outcomes, and the random effects

corresponding to type m. For the logit model, $P_{ijt}(\alpha_m)$ corresponds to equation 3 for periods

after 1998, and equation 4 for year 1998.

Step 4: Diagnose $E[r_{it}^j(\alpha_i)] = 0$

The following are the individual type-specific mean generalized residuals for drug j by type

$$m0_{ij}(\alpha_m) = \frac{\sum_{t=1999}^{2007} r_{it}^j(\alpha_m)}{10}$$

I estimate the mean generalized residual using the conditional probabilities as weights to

calculate the weighted average

$$m0_{ij} = \sum_{m=1}^{M} w_i^m * m0_{ij}(\alpha_m)$$

Then I take the mean over people of these residuals and report them along with the standard error

to test the null hypothesis that they are zero.

Step 5: Diagnose $E\left[r_{it}^j(\alpha_i), r_{it}^{j'}(\alpha_i)\right] = 0$

The type-specific correlation of generalized residuals is estimated as follows:

$$t0_i^{j,j'}(\alpha_m) = \frac{\sum_{t=1998}^{2007} r_{it}^j(\alpha_m) * r_{it}^{j'}(\alpha_m)}{10}$$

I use the conditional probabilities as weights to calculate the weighted average

$$t0_i^{j,j'} = \sum_{m=1}^{M} w_i^m * t0_i^{j,j'}(\alpha_m)$$

Then I take the mean over people of these residuals and report them along with the standard error

to test the null hypothesis that they are zero. In a similar way, I find 2nd-5th order autocorrelation

and cross-product correlations.

Appendix B: Classification error

My study was restricted to respondents from the NLSY97 who were not lost due to attrition for the entire 10 waves, and who answered questions related to past year alcohol, marijuana and cocaine consumption for all 10 waves as well. Three natural concerns arise from these restrictions: (1) Are respondents being lost due to attrition randomly? (2) Are respondents purposely avoiding responding to the drug-related questions? (3) Are respondents answering the questions truthfully, and, if not, to what extent does this misreporting affect the estimated parameters? This section will address these three concerns.

Is attrition random?

The first column of Table A1 reports mean summary statistics for the entire sample of 8984 respondents while the second column reports mean summary statistics for the subsample of respondents who were not lost due to attrition between 1997 and 2007, regardless of whether they avoided the drug questions. The first wave on NLSY97 started with 8984 respondents, and 62% (N=5623) of respondents completed all surveys from 1997 to 2007.

Columns 1 and 2 of Table A1 report that the average age at first wave, the share of respondents who have a father present in the household at the first wave, and the age of substance initiation, as well as the probability of ever consuming alcohol, marijuana or cocaine, remain almost unchanged when I restrict the sample to respondents who were not lost due to attrition between 1997 and 2007.

Did respondents purposely avoid drug-related questions?

The analysis reported in this paper is restricted to respondents for whom I have non-missing answers to their alcohol, marijuana, and cocaine-related question for all 10 waves. That implies that they were not lost due to attrition and also did not purposely avoid the drug-related questions. The mean summary statistics for this subsample of 5112 respondents is reported in column 3. After dropping observations with missing relevant covariates, I have 5108 respondents remaining. The entire paper is based on this subsample.

After restricting the sample to those who were not lost due to attrition, 94.59%, 92.78% and 94.47% of those respondents answered the alcohol, marijuana, and cocaine-related questions for all 10 waves, respectively.[71] This is evidence that respondents for whom I do not have the entire sequence of drug-related questions for all three drugs were lost mainly due to attrition instead of just avoiding the drug-related questions.

Misclassification Error:

Brener et al (2003) provides a detailed discussion of how the validity of self-reported risky behaviors data is affected by cognitive and situational factors. Because biochemical measures of substance use, such as breath tests to capture alcohol use or urinalysis to capture marijuana use, can only capture very recent substance use,[72] researchers need to rely on self-reported substance use data.

According to Brener et al. (2003), the most common sources of misreporting are: (1) Respondents have trouble remembering the exact time period in which a particular risky behavior occurred. In response to this concern, the main part of my analysis is based on reports

---

[71] Column 4, 5 and 6 report the mean summary statistics for respondents for whom I have non-missing alcohol, marijuana, or cocaine-related data for all 10 waves, respectively. Column 3 is restricted to those who reported non-missing alcohol, marijuana, and cocaine outcomes for all waves.

[72] The breath test captures alcohol use only within the 24 hours preceding the test; blood tests identify only heavy alcohol use; and urinalysis to detect recent marijuana use has a significant number of false negatives (Brener et. Al 2003).

of substance use, regardless of the amount, in the last year; this is easier to remember and hence more reliable than frequency or intensity of use. (2) The method of survey also affects the validity of self-reported measures of risky behaviors. For instance, CASI (or computer assisted self-interviewing) produces more accurate results than SAQ (pen-and-pencil self-administered questionnaire), and SAQ produces more measures of self-reported substance use than IAQ (in-person interview or interviewer-administered questionnaire). While the NLSY97 interviews were conducted using a Computer-Assisted Personal Interview (CAPI[73]), answers to sensitive questions were collected using audio computer-assisted self-interview (ACASI). The ACASI interviewing system enabled the respondent either to listen to the questions directly using headphones, or to read the questions from the laptop screen. This allows greater privacy.

Previous literature highlights the importance of accounting for classification error in dynamic discrete choice models for self-reported data on employment status (Keane and Sauer, 2009), full-time versus part-time school attendance, wages (Keane and Wolpin 2001, 2006), job change (Hausman et al. 1998) and many other outcomes.

Keane and Sauer (2010) and Hausman et al. (1998) provide a detailed discussion of how to incorporate, identify and estimate biased or unbiased classification error. Because the model is already computationally burdensome, and because misclassification can arise in alcohol, marijuana, and cocaine outcomes, incorporating misclassification error may be computationally infeasible, or at least extremely time consuming.

If respondents are still misreporting data, the most likely underreported outcome would be cocaine use. While it is obvious that the cocaine state dependence parameter would be biased

---

[73] Interviews conducted using a CAPI system were administered in person by an interviewer with a laptop computer. For more information, see http://www.bls.gov/nls/97guide/rd5/nls97ug2.pdf. A computerized interviewing system had a clear advantage at preventing invalid answers or answers that conflict with answers provided in previous waves, reducing inconsistencies across waves.

downward if cocaine use was underreported, I quantify the extent to which the estimated parameters change in the presence of classification error, by conducting the following simulation exercise:

(1) I estimate a logit model with seven mass points (Model A), and I call the vector of parameters $\{\theta_{m=7}\}$.

(2) I generate data using the parameters $\{\theta_{m=7}\}$. This newly generated data would be the true data if $\{\theta_{m=7}\}$ were the true parameters in the underlying process.

(3) I simulate a dataset where there is measurement error. I assume the dataset simulated in step 2 is the true data and $\{\theta_{m=7}\}$ are the true parameters. I create a new dataset where, every year, 20% of the respondents who truly consumed cocaine randomly report not having consumed any cocaine.

(4) I re-estimate the parameters using the data I created with measurement error in step 3, and I call these parameters $\{\theta_{m=7}^{error}\}$

(5) The new parameters $\{\theta_{m=7}^{error}\}$ tells me how much the true parameters $\{\theta_{m=7}\}$ would change if, independently in every year, 20% of random respondents who consumed cocaine reported not having consumed cocaine.

The estimated $\{\theta_{m=7}\}$ and $\{\theta_{m=7}^{error}\}$ reported in Table A2 indicate that introducing classification error underestimates cocaine state dependence by a statistically significant magnitude.[74]

Appendix C: Comparing data across datasets

Comparing the NLSY97 with other datasets is challenging because different datasets are collected using different timeframes and survey respondents in different age groups.

---

[74] I also performed the same analysis using a model with only 3 mass points to see whether models with a different number of mass points are more sensitive to the introduction of classification error. I get the same result; the only parameter that is affected is a statistically significant decrease in cocaine state dependence.

The 2010 report *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings*[75] compares measures of lifetime and recent drug use patterns for the following datasets, separately for teenagers (12-17 years old) and for young adults (18-25 years old), for years 2002 to 2010.

(1) National Survey on Drug Use and Health (NSDUH): Annual survey of civilians 12 years old and older who are not institutionalized, conducted in their household. The dataset collection began in 1971 using face-to-face-interviews. The interview system switched to a computer assisted system (CA) in 1998, in which the sensitive questions were administered via audio computer-assisted self-interviewing (ACASI), and the less sensitive questions were administered using computer-assisted personal interviewing (CAPI).

(2) Monitoring the Future (MTF): An ongoing annual survey of 8th, 10th and 12th graders, in which they complete a self-administered computerized interview in school, during a regular class period. A sample of each graduating class is selected to receive annual follow up questionnaires.

Table A3 compares the percentage of young adults (18 to 25 years old) who reported having used alcohol, marijuana, cocaine or other drugs in their lifetime, in the last year and in the last month in 2002.

For this comparison exercise, I focus on young adults as opposed to teenagers. If I had to compare teenagers 12 to 17, the best I could do would be comparing teenagers who were 12-17 in 1997 (NLSY97) with teenagers from the NSDUH in 2002. On the other hand, young adults

[75] Substance Abuse and Mental Health Services Administration, *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings*, NSDUH Series H-41, HHS Publication No. (SMA) 11-4658. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2011.

(18-23) at the NLSY97 in 2002 can be compared with young adults (18-25) in NSDUH and MTF (18-24) for the same year[76].

Measures of cocaine use from the NLSY97 are not directly comparable with measures of the NSDUH or MTF, because the NLSY97 groups cocaine with other hard drugs. The measures of lifetime alcohol and marijuana use are similar between the NLSY97 and the NSDUH, while the MTF reports slightly higher measures. A reverse pattern holds for measures of lifetime cocaine use, where the MTF reports lower prevalence than the NSDUH. These differences between the NSDUH and the MTF can be reconciled if respondents report higher prevalence of alcohol and marijuana when the interview is conducted at school (MTF) than at home (NSDUH) due to perceived differences in privacy. On the other hand, the share of lifetime cocaine consumption reported by the MTF would have probably been higher if it had included students who were absent from school the day of the interview or had dropped out.

Measures of past year alcohol and marijuana consumption also followed a similar pattern where the NLSY97 lead to a lower estimate than the NSDUH, and the NSDUH had a lower estimate than the MTF. This can be reconciled by the NSDUH and MTF containing a higher mean average age in 2002 than the NLSY97, and by the fact that students may have more privacy in school than in the household. Measures of cocaine consumption in the last year are similar between NSDUH and MTF, while I cannot compare that reported by the NLSY97, because it includes other hard drugs.

Measures of alcohol use in the last month follow a similar pattern as measures of alcohol use in the last year. The NLSY97 does not include a measure of cocaine use in the last month, as

---

[76] During years 2002 and after, the sensitive questions in the NLSY97 and NSDUH were asked using the same ACASI system, which makes these reports directly comparable.

the NLSY97 only reports whether the respondent consumed cocaine at all in the last year and how many times he or she used it in the last year.

Appendix D: Inclusion of Time-Invariant and Time-Varying Covariates

Model A includes a very limited list of time-invariant covariates, which is not problematic under the assumption that the unobserved heterogeneity will capture most of the permanent characteristics that affect individual preferences for a particular drug. In this section, I estimate Model A with only 3 mass points[77] using equations 1 as the utility of consuming drug j in year t and equation 2 as the initial conditions equation. For this section, I leave the initial conditions equation unchanged, and explore the model with additional covariates in equation 1.

Columns 1,4 and 7 report parameter estimates corresponding to the latent utility (equation 1) for alcohol. Columns 2,5, and 8 correspond to marijuana, and finally columns 3,6, and 9 correspond to cocaine.

Panel A of Table A4 shows that the results remain robust to the inclusion of time-invariant characteristics. The first three columns show the state dependence and stepping stone parameters for Model A with only 3 mass points with no additional covariates.[78] Columns 4, 5 and 6 show that the results remain robust to the inclusion of indicators for race in the latent utilities (an indicator for black and one for Hispanic). Finally, columns 7, 8 and 9 show that the results are also robust to the inclusion of an indicator for whether the respondent was originally from an urban area. In comparing models that include different time-invariant covariates, I

---

[77] The benchmark model (Model A) was estimated with seven mass points. The model I present for this section only has three mass points. While this model is not the preferred specification, the point of this section is to show that these models are robust to the inclusion of time-invariant and time-varying demographics.
[78] The benchmark model (Model A) included two time-invariant characteristics (male and an indicator for whether the respondent lived in a single parent household in the first wave) and one time-varying characteristic (indicators for different age categories). While the benchmark model includes seven mass points, I estimate this model with only three mass points for simplicity.

observe that once I include geographic information,[79] one of the mass points for hard drugs tends to negative infinity (logit parameters of -13, -17 for marijuana and cocaine). This means that there is a group of individuals who never consume hard drugs. While the mass points change to the inclusion of time-invariant characteristics, the state dependence and stepping stone parameters remain robust.

Panel B of Table A4 reports the state dependence and stepping stone parameters after the inclusion of time-varying covariates. Because college attendance may create a more drug friendly environment, I expect college attendance to influence the latent preference for consuming alcohol, marijuana, and hard drugs. For instance, if college attendance in time t increases the propensity to consume alcohol in time t with no direct effect on alcohol consumption at t+1, omitting college attendance could bias the state dependence and stepping stone parameters. The first three columns report that the state dependence and stepping stone parameters remain robust after the inclusion of time-varying indicators for enrollment in a two-year college, enrollment in a four-year college, and employment. Similarly, columns 4, 5 and 6 report that the estimates are robust to time-varying region of residence.[80] Finally, the last three columns indicate that the estimates are also robust to the inclusion of state and year specific unemployment rate.

I allowed all these time-invariant and time-varying coefficients to be drug specific.[81]

---

[79] I also estimated this model with the inclusion of indicators for the original region of residence for the respondent instead of an indicator for whether the respondent lived originally in an urban area. The results were very similar.
[80] Among 4998 respondents with non-missing regions for all 10 years, 4261 of them never changed regions in the entire ten waves.
[81] Being currently enrolled in college (whether it is a two-year or four-year college) and being currently employed increase the preference for alcohol and decrease the preference for hard drugs. The effects of college enrollment on marijuana are heterogeneous. The preference for marijuana is lower when the respondent is enrolled in a two-year college, increases when enrolled in a four-year college, and decreases when the respondent is employed. These tables can be provided upon request.

Acknowledgements

REFERENCES

Anderson, M. and D. Rees, 2012, Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption. IZA Discussion Paper No. 6112. Available at SSRN: http://ssrn.com/abstract=1965129

Altonji, J., S. Cattan and I. Ware, 2010, Identifying sibling influence on teenage substance use. National Bureau of Economics Research Working Paper 16508.

Andrews, D.W.K.,1988, Chi square diagnostic tests for econometric models: Theory. Econometrica 56, 1419-1453.

Arcidiacono, P., A. Khwaja, and L. Ouyang, 2011, Habit persistence and teen sex: Could increased access to contraception have unintended consequences for teen pregnancies? Journal of Business and Economics Statistics.

Baker. M. and A. Melino, 2000, Duration dependence and nonparametric heterogeneity: A Monte Carlo Study. Journal of Econometrics 96(2), 357-393.

Becker, G. and K. Murphy, 1988, A theory of rational addiction. The Journal of Political Economy, Vol 96 (4); 675-700

Brown, G. and L. Silverman, 1974, The retail price of heroin: Estimation and applications. Journal of the American Statistical Association 347 (69), 595-606.

Becker, G., M. Grossman and K. Murphy, 1991, Rational Addiction and the Effect of Price on Consumption. The American Economic Review Papers and Proceedings of the Hundred and Third Annual Meeting of the American Economic Association 81(2), 237-241.

Brener, N., J. Billy and W. Grady, 2003, Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence from the Scientific Literature. Journal of Health Adolescent Health 33, 436-457

Cameron, A. and P. Trivedi, 2005, Microeconometrics: methods and Applications. Cambridge University Press.

Card, D. and D. Hyslop, 2005, Estimating the effects of a time-limited earnings subsidy for welfare leavers. Econometrica 73, 1723-1770.

Carpenter, C., and C. Dobkin, 2009, The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age. American Economic Journal: Applied Economics 1(1), 164–82.

Carpenter, C., L. Johnston, D. Kloska and P. O'Malley, 2007, Alcohol Control Policies and Youth Alcohol Consumption: Evidence from 28 years of monitoring the future. Journal of Economic Analysis and Policy Vol 7(1), article 25, 1-21.

Carpenter, C. ,2004a, How Do Zero Tolerance Drunk Driving Laws Work? Journal of Health Economics 23(1), 61-83.

Carpenter, C. , 2004b, Heavy Alcohol Use and Youth Suicide: Evidence from Tougher Drunk Driving Laws. *Journal of Policy Analysis and Management* 23(4), 831-842.

Carpenter, C., 2008, Heavy Alcohol Use and Crime: Evidence from Underage Drunk Driving Laws. Journal of Law and Economics 50(3), 539-557.

Carpenter, C. and C. Dobkin, 2010, The Drinking Age, Alcohol Consumption, and Crime. Working Paper, University of California Irvine

Cawley, J., 2008, *Reefer Madness*, Frank the Tank or *Pretty Woman*: To What Extent Do Addictive Behaviors Respond to Incentives?, in Frank Sloan and Hirschel Kasper (editors), Incentives and Choice in Health and Health Care, (MIT Press: Cambridge, MA).

Chaloupka, F., R. Pacula, M. Farrelly, L. Johnston, and P.M. O'Malley, 1999, Do higher cigarette prices encourage youth to use marijuana? National Bureau of Economic Research, Working Paper No. 6939

Chaloupka, F., M. Grossman, and J. Tauras, 1999, The demand for cocaine and marijuana by youth. In The Economic Analysis of Substance Use and Abuse: An integration of Econometrics and Behavioral Economic Research. Frank J. Chaloupka, Michael Grossman, Warren K. Bickel and Henry Saffer, editors

Chesher, A. and M. Irish, 1987, Residual analysis in the grouped and censored normal linear model. Journal of Econometrics 34, 33-61.

Dee, T., 1999, State alcohol policies, teen drinking and traffic fatalities. Journal of Public Economics 72(2), 289-315.

Dhaval, Dave, 2006, The effects of cocaine and heroin price on drug-related emergency department visits. Journal of health economics 25 (2), 311-333.

DeSimone, J., 1998, Is marijuana a gateway drug? Eastern Economic Journal 24(2), 149-164.

DiNardo, J. and T. Lemieux, 2001, Alcohol, marijuana, and American youth: the unintended consequences of government regulation. Journal of Health Economics, 20(6), 991-1010

Dobkin, C. and N. Nicosia, 2009, The war on drugs: methamphetamine, public health and crime. American Economic Review 99(1), 324-349.

Enberg, J., P. Gottschalk, and D. Wolf , 1990, A random-effects logit model of work-welfare transitions. Journal of Econometrics 43(1-2), 63-75.

Gentzkow, M., 2007, Valuing new goods in a model with complementarity: Online newspapers. American Economic Review 97(3).

Gourieroux, C, A. Monfort, E. Renault, and A. Trognon , 1987, Generalized residuals. Journal of Econometrics 34(1-2), 5-32.

Grossman, M. and F. Chaloupka, 1998, The demand for cocaine by young adults: A rational addiction approach. Journal of Health Economics 17(4), 427-474.

Halliday, T., 2008, Heterogeneity, state dependence and health. Econometrics Journal, 11(3), 499-516.

Ham, J. and R. Lalonde ,1996, The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. Econometrica 64(1), 175-205.

Hausman, J. A., J. Abrevaya, and F. Scott-Morton , 1998, Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics 87, 239-269

Heckman, J., 1979, Statistical models for discrete panel data, in C. F. Manski and D.

McFadden(Eds.), Structural Analysis of Discrete Data with Econometric Applications.

Cambridge, Mass.: MIT Press, 1981, pp. 114-78

Heckman, J., 1981a, Heterogeneity and state dependence, in S. Rosen: Studies in Labor Markets.

Chicago: University of Chicago Press, 1981, pp. 91-140

Heckman, J., 1981b, The incidental parameters problem and the problem of initial conditions in

estimating a discrete time-discrete data stochastic process, in C. F. Manski and D. McFadden:

Structural Analysis of Discrete Data with Econometric Applications. Cambridge, Mass.: MIT

Press, 1981, pp. 179-95

Heckman, J. and S. Navarro, 2007 , Dynamic discrete choice and dynamic treatment effects.

Journal of Econometrics 136, 341-396.

Heckman, J. and B. Singer, 1984, A method for minimizing the impact of distributional

assumptions in econometric models for duration data. Econometrica 52(2), 271-320.

Hoyt, G. and F. Chaloupka, 1994, Effects of survey conditions of self-reported substance use.

Contemporary Economic Policy 12(3), 109-121.

Hyslop, D., 1999, State dependence, serial correlation and heterogeneity in intertemporal labor

force participation of married women. Econometrica 67(6), 1255-1294.

Johnston, L. D., P. O'Malley, J. Bachman, and J. Schulenberg, 2011, Monitoring the Future national survey results on drug use, 1975-2009; Vol I. Secondary school students (NIH Publication 10-7584). Bethesda. MD: National Institute on Drug Abuse.

Keane, M., 1997, Modeling heterogeneity and state dependence in consumer choice behavior. Journal of Business & Economic Statistics 15(3), 310-327.

Keane, M. and K.Wolpin, 2006, The role of labor and marriage markets, preference heterogeneity and the welfare system in the life cycle decisions of black, hispanic and white women. PIER Working paper N. 06-004.

Keane, M. and R.Sauer, 2009, Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior. Econometrica 77 (3), 975-991.

Keane, M. and R. Sauer, 2010, A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables. International Economic Review 51(4), 925-958.

Kenkel, D. A. Mathios, and R. Pacula , 2001, Economics of youth drug use, addiction and gateway effects. Addiction 96(1), 151-164.

Lynskey, M., J. Vink, and D. Boomsma, 2006, Early Onset Cannabis Use and Progression to Other Drug Use in a Sample of Dutch Twins., Behavior Genetics 36(2).

Markowitz, Sara, 2000.The price of alcohol, wife abuse and husband abuse. Southern Economic Joutnsl 67(2), 279-303.

McCoun. R. , 1998, In what sense (if any) is marijuana a gateway drug. FAS Drug Policy Analysis Bulletin 4, 3-5.

McCall, B., 1994, Specification diagnostics for duration models: A martingale approach. Journal of econometrics 60(1-2), 293-312.

Meer, Jonathan (forthcoming), The habit of giving. Economic Inquiry.

Mills, C. and H. Noyes, 1984, Patterns and correlates of initial and subsequent drug use among adolescents. Journal of Consulting and Clinical Psychology 52(2).

Moore D., 1977, Generalized inverses, Wald's method, and the construction of Chi-square tests of fit. Journal of the American Statistical Association, 131-137.

Moore, W., S. Pedlow, P. Krishnamurty, K. Wolter, and I. Chicago, 2000, National longitudinal survey of youth 1997 (NLSY97).

Newcomb, M. and P. Bentler, 1986, Cocaine use among adolescents, Longitudinal associations with social context, psychopathology, and use of other substances. Addictive bahaviors, 11(3), 263-273.

Office of National Drug Control Policy, 1997, An Empirical Examination of Counterdrug Interdiction Program Effectiveness. Institute for Defense Analysis (IDA) Paper P-3219.

Office of National Drug Control Policy, 2008, The Price and Purity of Illicit Drugs: 1981-2007. Institute for Defense Analysis (IDA) Paper P-4369, electronically accessible through the following World Wide Web address http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/bullet_1.pdf . The accompanying *Technical Report* is available at http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/price_purity_tech_rpt07.pdf .

Orphanides, A. and D. Zervos, 1995, Rational addiction with learning and regret. Journal of Political Economy 103(4), 739-758.

Pacula, R., 1998, Adolescent alcohol and marijuana consumption: Is there really a gateway effect. National Bureau of Economic Research Working Paper 6348

Pacula, R., M. Grossman, F. Chaloupka, P. M. O'Malley L.D. Johnson, and M.C. Farrelly , 2001, Marijuana and youth, in J. Gruber (Eds.), Risky Behavior among Youths: An Economic Analysis Chapter 6, Chicago : University of Chicago Press, p. 271-326

Plant. M, 1984, An empirical analysis of welfare dependence. The American Economic Review, 673-684.

Pollack, R., 1970, Habit Formation and Dynamic Demand Functions. Journal of Political Economy 78(4),745-763.

Pollack, R., 1976, Habit Formation and Long Run Utility Functions. Journal of Economic Theory 13, 272-297.

Prinz. A. , 1997, Do European drugs policies mater? Economic Policy 12(25), 373-385.

Pudney. S., 2003, The road to ruin? Sequences of initiation to drugs and crime in Britain. Economic Journal, 182-198.

Register, C., D. Williams, and P. Grimes, 2001, Adolescent Drug Use and Educational Attainement. Education Economics 9(1) , 1-18.

Saffer, H. and Chaloupka, F., 1999. The demand for illicit drugs. Economic Inquiry, Western Economic Association International 37(3), 401-411.

Silverman, L. and N. Spruill, 1977, Urban Crime and the Price of Heroin, Journal of Urban Economics.

Stigler, G. and G. Becker, 1977, De gustibus non est disputandum. American Economic Review 67(2), 76-90.

Substance Abuse and Mental Health Services Administration, Office of Applied Studies. *Treatment Episode Data Set (TEDS): 1992-2001. National Admissions to Substance Abuse Treatment Services*, DASIS Series: S-20, DHHS Publication No. (SMA) 03-3778, Rockville, MD, 2003. http://wwwdasis.samhsa.gov/teds01/TEDS2K1Index.htm.

Substance Abuse and Mental Health Services Administration, 2011, *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings*, NSDUH Series H-41, HHS Publication No. (SMA) 11-4658. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Taubman, P., 1991, Externalities and decriminalization of drugs. In Drug policy in the United States, ed. Melvyn B. Krauss and Edward P. Lazear. Stanford, California: Hoover Institution Press.

Train, K. (2003) Discrete choice methods with simulation. Cambridge University Press.

Van Ours J., 2003, Is cannabis a stepping-stone for cocaine? Journal of Health Economics 22(4), 539-554.

Van Ours, J., 2006, Dynamics in the use of drugs. Health economics, 15(12),1283-1294.

Van Ours, J., 2007, Cannabis use when it's legal. Addictive behaviors 32(7),1441-1450.

Van Ours, J. and J. Williams, 2009, Why parents worry: Initiation into cannabis use by youth and their educational attainment. Journal of Health Economics 28(1), 132-142.

Wooldridge, J., 2005, Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics 20(1), 39-54.

Yamada, T., M. Kendix, and T. Yamada, 1996, The impact of alcohol consumption and marijuana use on high school graduation. Health Economics 5(1), 77-92.

Yamaguchi K. and D.B. Kandel, 1984, Patterns of drug use from adolescence to young adulthood: Sequences of progression. American Journal of Public Health 74(7), 668-672.

Yoruk, B. and C. Yoruk, 2011, The impact of minimum legal drinking age laws on alcohol consumption, smoking, and marijuana use: Evidence from a regression discontinuity design using exact date of birth. Journal of Health Economics 2011 (30), 740-752

Yu, J. and W. Williford, 1992, The Age of Alcohol Onset and Alcohol, Cigarette, and Marijuana Use Patterns: An Analysis of Drug Use Progression of Young Adults in New York State. The International Journal of the Addictions 27(11), 1313-1323.

$alcohol_t \rightarrow \{P(alcohol_{t+1}) = 87.75\%$

$No\ alcohol_t \rightarrow \{P(alcohol_{t+1}) = 31.73\%$

$marijuana_t \rightarrow \{P(marijuana_{t+1}) = 66.99\%$

$No\ marijuana_t \rightarrow \{P(marijuana_{t+1}) = 9.09\%$

$cocaine_t \rightarrow \{P(cocaine_{t+1}) = 43.40\%$

$No\ cocaine_t \rightarrow \{P(cocaine_{t+1}) = 3.48\%$
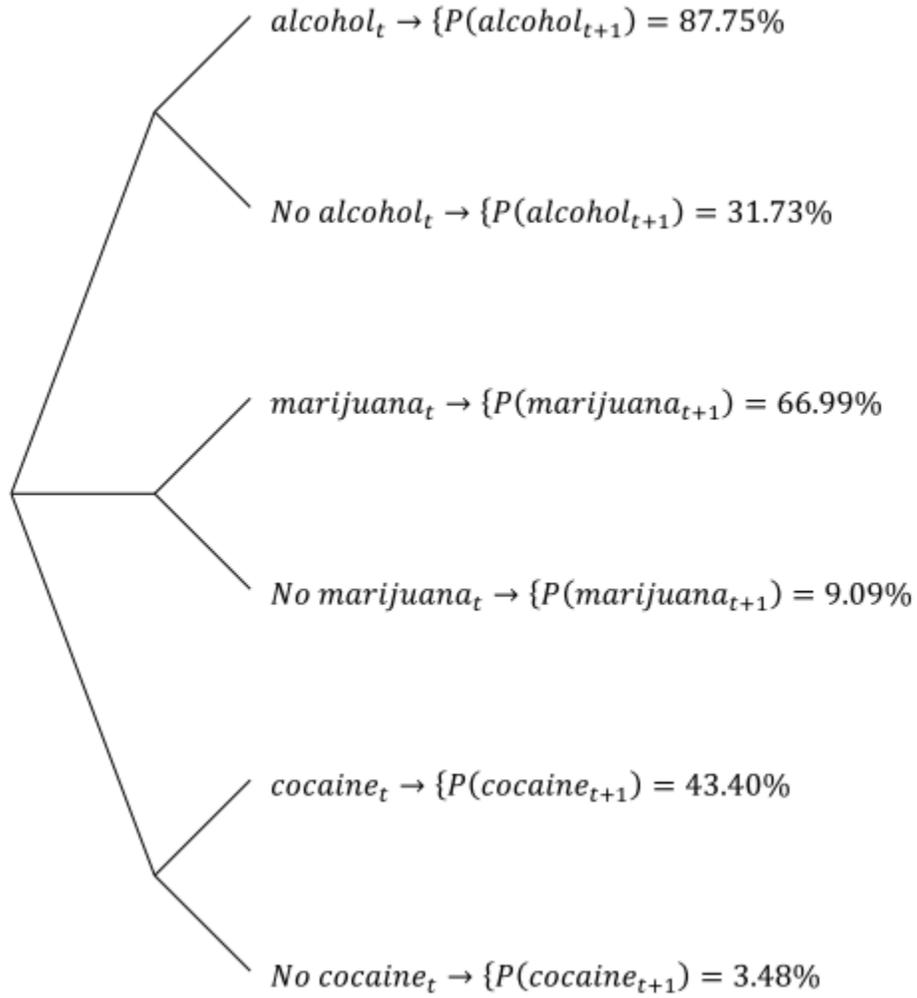
Figure 1: Drug Persistence

Note: I compute these probabilities at the yearly level, starting in 1999, and then took the average

$$alcohol_t \rightarrow \begin{cases} P(marijuana_{t+1}) = 22\% \begin{pmatrix} 71\% \; if \; mar_t = 1 \\ 7\% \; if \; mar_t = 0 \end{pmatrix} \\ \\ P(cocaine_{t+1}) = 8.5\% \begin{pmatrix} 44\% \; if \; coc_t = 1 \\ 5\% \; if \; coc_t = 0 \end{pmatrix} \end{cases}$$

$$No \; alcohol_t \rightarrow \begin{cases} P(marijuana_{t+1}) = 4\% \begin{pmatrix} 46\% \; if \; mar_t = 1 \\ 3\% \; if \; mar_t = 0 \end{pmatrix} \\ \\ P(cocaine_{t+1}) = 1.3\% \begin{pmatrix} 14\% \; if \; coc_t = 1 \\ 1.2\% \; if \; coc_t = 0 \end{pmatrix} \end{cases}$$

$$marijuana_t \rightarrow P(cocaine_{t+1}) = 18.9\% \begin{pmatrix} 47\% \; if \; coc_t = 1 \\ 11\% \; if \; coc_t = 0 \end{pmatrix}$$

$$No \; marijuana_t \rightarrow P(cocaine_{t+1}) = 2.1\% \begin{pmatrix} 27\% \; if \; coc_t = 1 \\ 1.8\% \; if \; coc_t = 0 \end{pmatrix}$$

Figure 2: "Transitions" from Softer to Harder Drugs

Note: I computed these probabilities for every year starting in 1999, and then took the average.
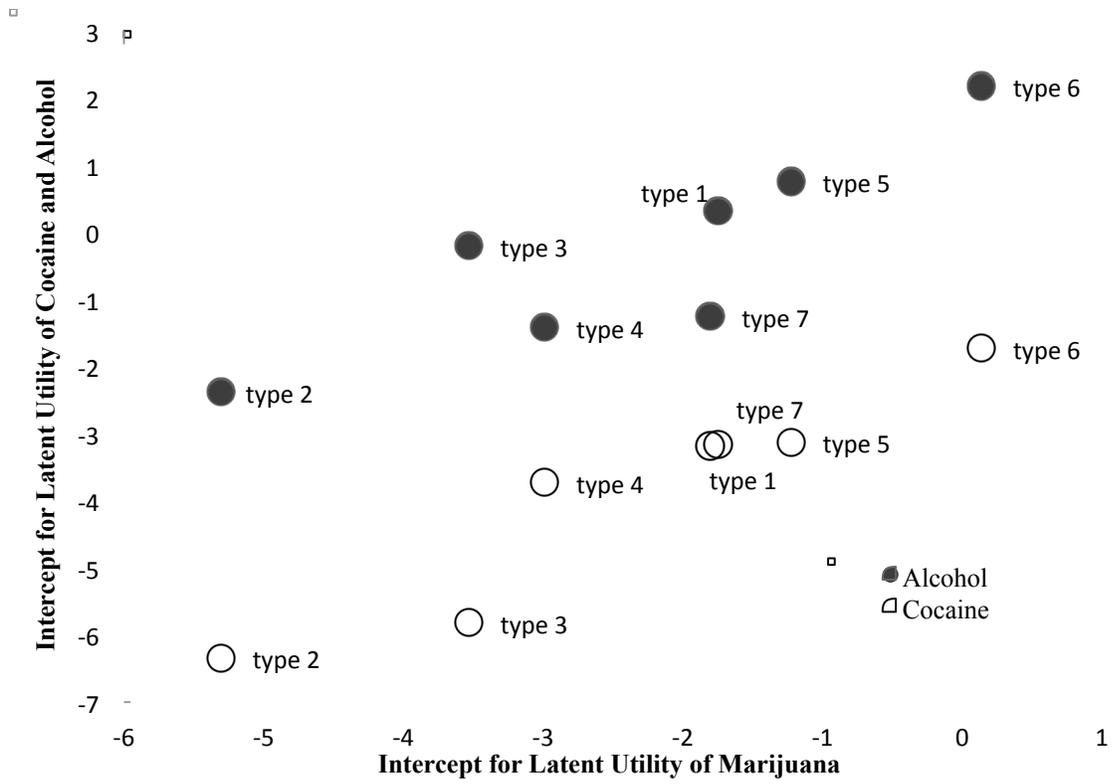
Figure 3A: Arbitrarily correlated unobserved heterogeneity (Model A)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.
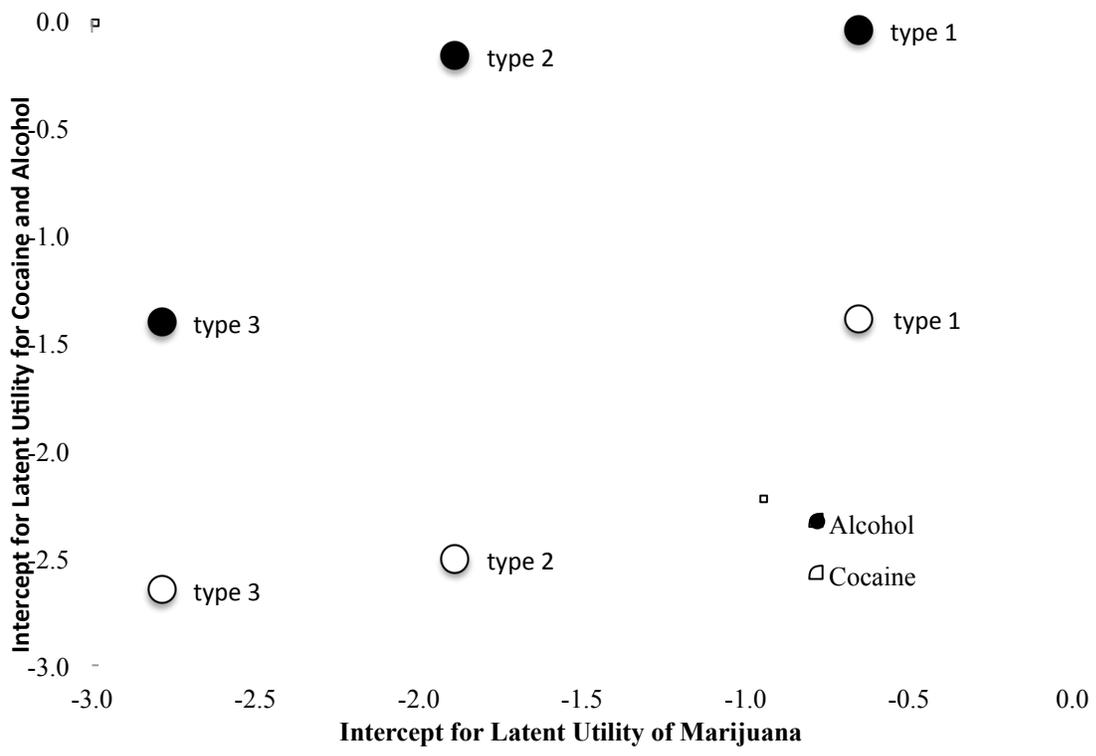
Figure 3B: Arbitrarily correlated unobserved heterogeneity (Model B)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.
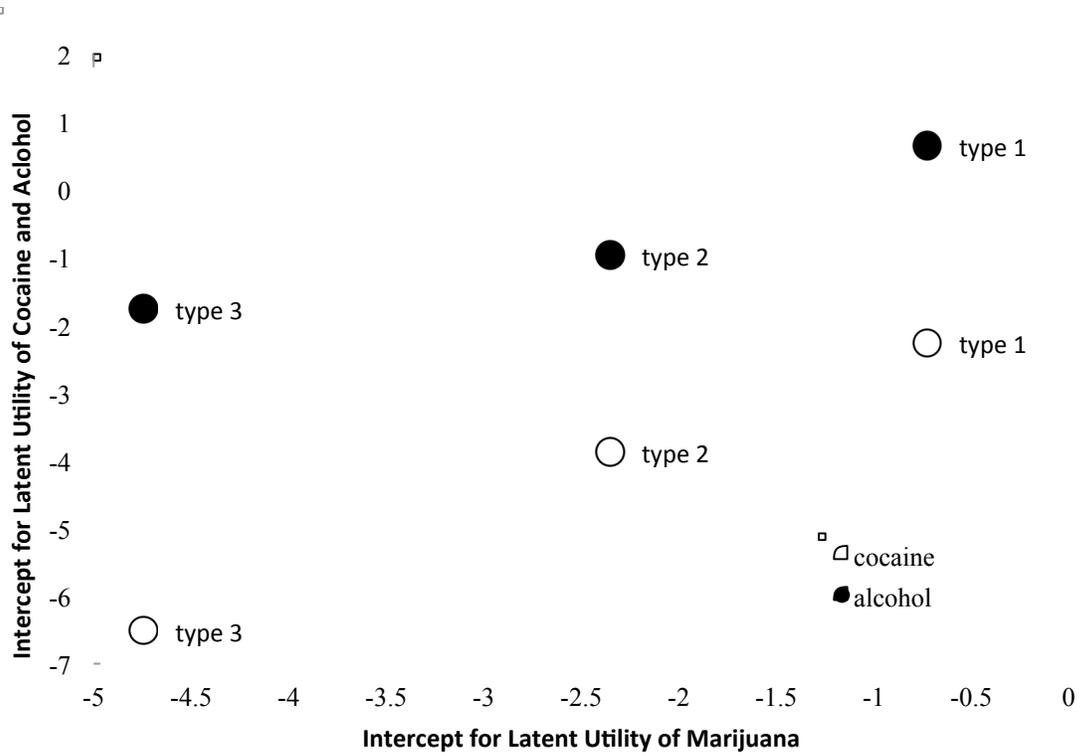
Figure 3C: Arbitrarily correlated unobserved heterogeneity (Model C)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.
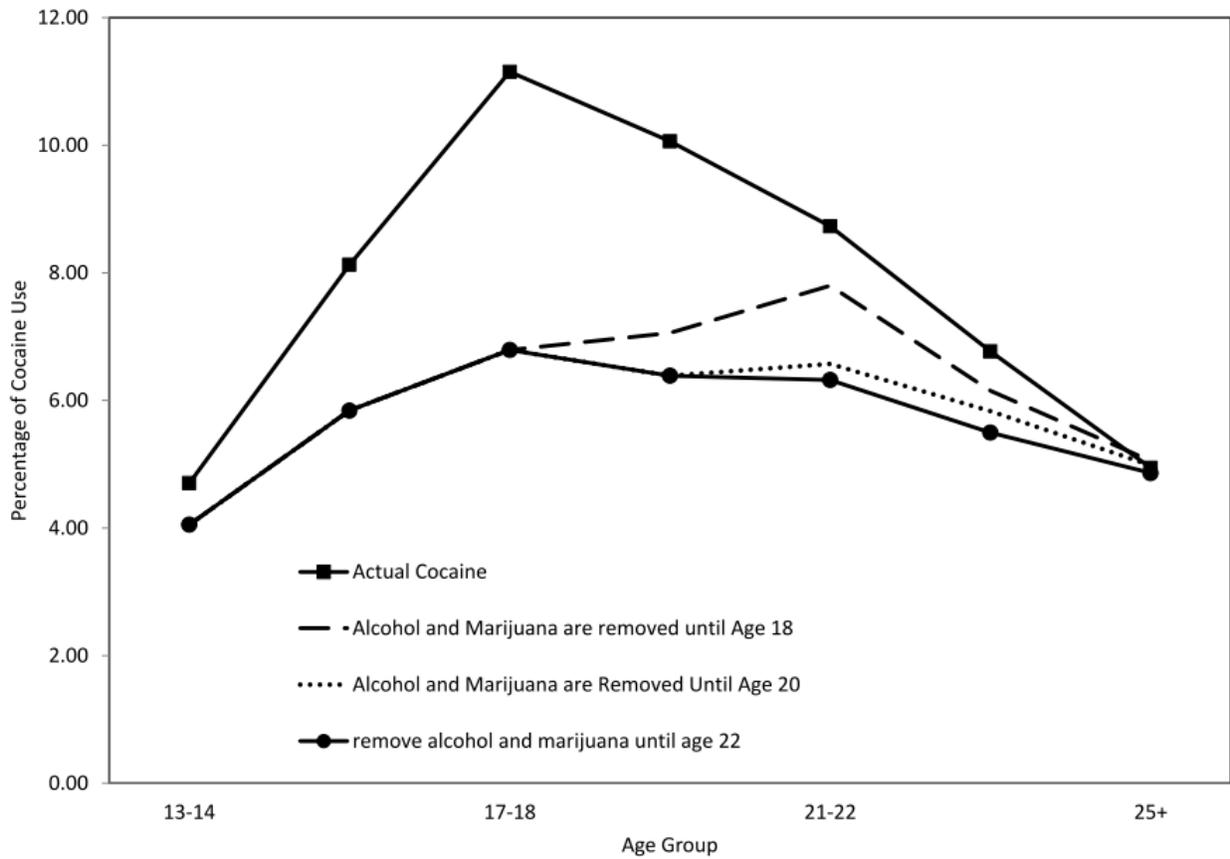
Figure 4: Simulated impacts of preventing drug use at early ages

Note: Age profile of cocaine consumption, and the simulated age profile under the assumption that alcohol and marijuana are completely removed at early ages.
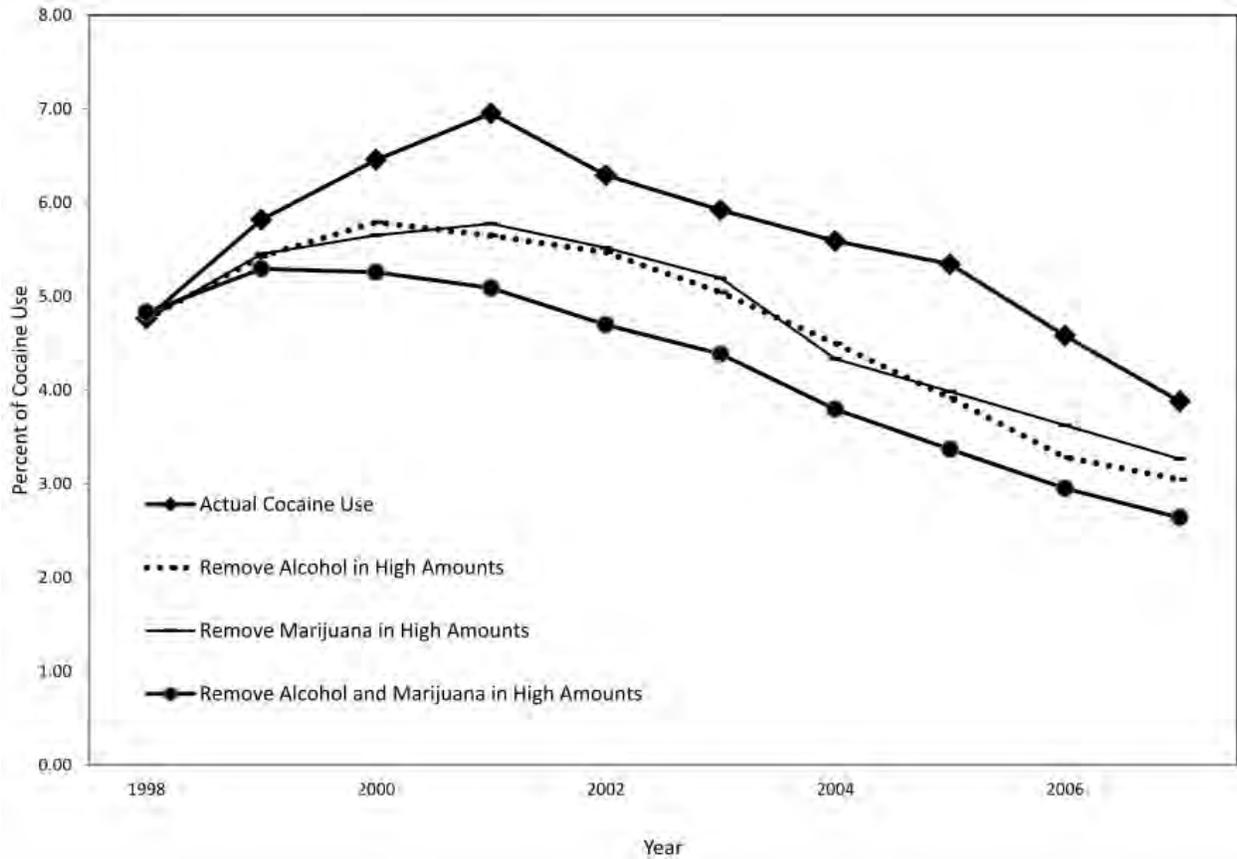
Figure 5: Simulated impacts of regulating intensity of drug use

Share of cocaine users by year and the simulated share after I replace simulated high levels of consumption of alcohol and marijuana with low levels of consumption.
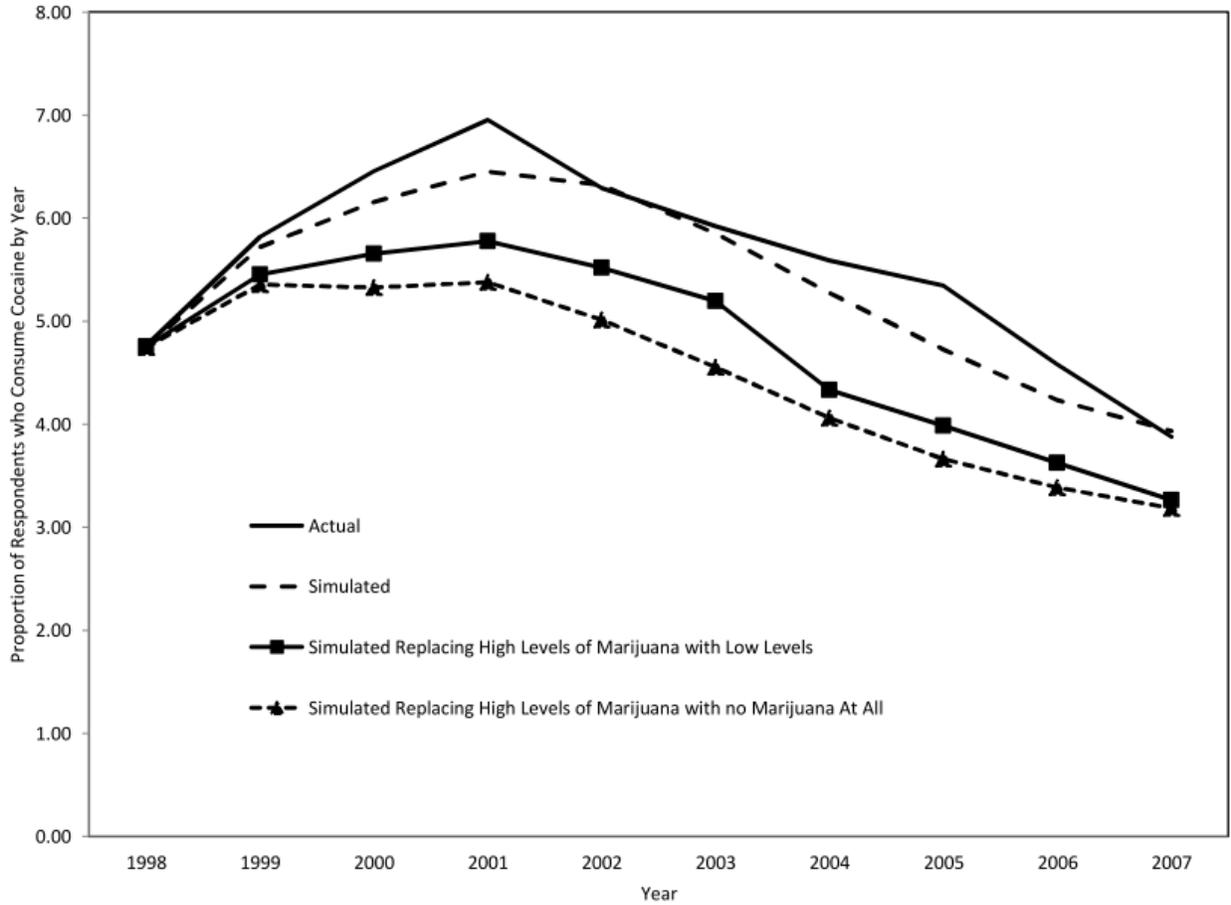
Figure 6: Simulated impacts of restrictions on marijuana intensity of use

Note: Share of cocaine users by year and the simulated share after I replace simulated high levels of consumption of marijuana with low levels of consumption, or not consumption at all.

Table 1: Summary Statistics

	Full Sample		Non-missing Drug Questions for all 10 waves		
				Ever Used	
			Alcohol	Mar	Cocaine
Panel A: Sample Mean					
Age(1997)	14.35	14.23	14.24	14.24	14.21
	(1.49)	(1.47)	(1.47)	(1.48)	(1.49)
Male	0.51	0.47	0.47	0.50	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Father in household	0.72	0.73	0.73	0.72	0.72
	(0.45)	(0.44)	(0.44)	(0.45)	(0.45)
Age First Drink	14.99	15.03	15.03	14.04	13.68
	(3.42)	(3.41)	(3.42)	(2.92)	(2.80)
Age First Marijuana	16.16	16.17	16.17	16.17	15.30
	(3.02)	(3.01)	(3.01)	(3.02)	(2.52)
Age First Cocaine	17.59	17.58	17.59	17.49	17.58
	(3.32)	(3.32)	(3.32)	(3.32)	(3.32)
Percentage of respondents with non-missing answers for drug-related questions(a)					
P(non-missing alcohol)	59.21	100.00	100.00	100.00	100.00
P(non-missing marijuana)	58.07	100.00	100.00	100.00	100.00
P(non-missing cocaine)	59.13	100.00	100.00	100.00	100.00
Percentage of respondents who reported having consumed drugs at least once (b)					
P(ever alcohol)	92.39	94.58	100.00	99.39	99.52
P(ever marijuana)	57.66	57.34	60.25	100.00	92.61
P(ever cocaine)	23.12	24.63	25.92	39.78	100.00
N	8984	5112	4835	2931	1259

Panel B: Order of Use Among Individuals that Used all Three Substances Eventually

	N	Percent
Alcohol<Marijuana<Cocaine	473	40.78
Alcohol<Cocaine<Marijuana	71	6.12
Marijuana<Cocaine<Alcohol	3	0.26
Marijuana<Alcohol<Cocaine	80	6.90
Cocaine<Alcohol<Marijuana	16	1.38
Cocaine<Marijuana<Alcohol	9	0.78
Alcohol<=Marijuana<=Cocaine	891	76.81
N		1160

Note: Standard errors are in parenthesis. These summary statistics are unweighted.
Column (1) reports demographics for the entire survey.
Column (2) reports demographics for the subsample for whom we have non-missing answers to the drug-related questions from 1997 to 2007.
Column (3)-(4)-(5) corresponds to the subsample of column 2 that consumed alcohol, marijuana, and cocaine, respectively.
P(non-missing drug j)=percentage of the sample for whom we have non-missing answers to questions related to drug j from 1997 to 2007.
P(Ever drug j)=percentage of the sample who consumed drug j at least once by 2007.
In Panel B, the row corresponding to "Alcohol<Marijuana<Cocaine" corresponds to the share of respondents who eventually consumed all three drugs but consumed alcohol before marijuana, and marijuana before cocaine. The symbol ≤ means "before or during the same year."

Table 2: Parameter Estimates for the Models with Homogeneous State Dependence and Stepping-Stone

	Model A: Logit (1st Order)			Model B: Probit (1st Order)			Model C: Logit (2nd Order)		
	Alcohol	Marijuana	Cocaine	Alcohol	Marijuana	Cocaine	Alcohol	Marijuana	Cocaine
Lag alcohol	1.74	0.28	0.33	1.19	0.45	0.34	1.89	0.19	0.22
	(0.03)	(0.05)	(0.09)	(0.04)	(0.06)	(0.09)	(0.04)	(0.05)	(0.10)
Lag mar	0.07	1.66	0.30	0.18	0.99	0.24	-0.15	1.58	0.35
	(0.05)	(0.04)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Lag coc	-0.08	-0.04	1.58	0.10	0.07	0.99	-0.50	-0.10	1.71
	(0.09)	(0.06)	(0.06)	(0.09)	(0.07)	(0.07)	(0.10)	(0.06)	(0.08)
$cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar})$					0.58				
$cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{coc})$					0.38				
$cov(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc})$					0.46				
Y_{t-2}							0.99	0.68	1.01
							(0.05)	(0.06)	(0.09)
$Y_{t-1} * Y_{t-2}$							0.13	-0.04	-0.56
							(0.06)	(0.08)	(0.12)
N. obs		5108			1000			5108	
LL		48024			9369			48133	
N params		87			62			107	
N. mass points		7			3			3	
AIC		96222			18862			96479	
BIC		96791			19166			97178	
PCGF	417	123	43	76	64	18	212	66	22

Note: Parameter estimates corresponding to the latent utility for drug j in period t, where j={alcohol, marijuana, cocaine}. See section 3 for a discussion of Model A, B, and C. The probit model estimates the following parameters $\phi_{12} = 1.311$ ($SE = 0.077$), $\phi_{13} = 0.803$ ($SE = 0.103$) and $\phi_{23} = 0.989$ ($SE = 0.072$). The following correlations are by construction between -1 and 1. $cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}) = -1 + 2 * \frac{\exp(\phi_{12})}{1+\exp(\phi_{12})}$, $cov(\varepsilon_{it}^{drink}, \varepsilon_{it}^{coc}) = -1 + 2 * \frac{\exp(\phi_{13})}{1+\exp(\phi_{13})}$, and $cov(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}) = -1 + 2 * \frac{\exp(\phi_{23})}{1+\exp(\phi_{23})}$

Table 3 : Marginal Effects

	Model A	Model B	Model C Second Lag=0	Second Lag=1
Panel A: State Dependence				
Alcohol	0.203	0.240	0.361	0.149
Marijuana	0.153	0.182	0.098	0.237
Cocaine	0.065	0.097	0.052	0.377
Panel B: Stepping Stone Effect				
From alcohol to marijuana	0.026	0.084	0.016	
From alcohol to cocaine	0.014	0.033	0.008	
From marijuana to cocaine	0.012	0.024	0.013	

Note: For models A and C (Logit), I compute $\frac{\partial P_{ijt}}{\partial z_j}$ for each individual i and period t, where $\frac{\partial P_{ijt}}{\partial z_j} = P_{ijt} * (1 - P_{ijt}) * \beta_j$. I compute the average from year 1999 to 2007 for Model A. For Model C, I compute $\frac{\partial P_{ijt}}{\partial z_j}$ for years 2000 to 2007 and take the average. Since Model C includes an interacted term of first and second-order state dependence, I compute the marginal effect of first-order state dependence separately for periods where the second-order lagged outcome was 0 and for periods where the second-order lagged outcome was 1. For Model B, I compute $\frac{\partial P_{ijt}}{\partial z_j} = \phi\left(V_{ijt}(\alpha_i)\right) * \beta_j = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}[V_{ijt}(\alpha_i)]^2\} * \beta_j$ for years 1999 to 2007, and take the average. The marginal effects are computed using simulated data because a random effect needs to be assigned to compute P_{ijt}. I repeat the process of computing $\frac{\partial P_{ijt}}{\partial z_j}$ ten times and take the average.

Table 4: Discrete Effect of Lagged Consumption on Current Consumption

	Model A: Logit (1st Order)			Model B: Probit (1st Order)			Model C: Logit (2nd Order)		
	Actual	Sim	Sim(γ_{kj}=0)	Actual	Sim	Sim(γ_{kj}=0)	Actual	Sim	Sim(γ_{kj}=0)
Panel A: Persistence of Use Within Drug (State Dependence)									
$P(Y_{it}^{drink}=1\|Y_{i,t-1}^{drink}=1)$	87.75	87.77	64.67	90.13	89.36	57.09	87.75	87.94	59.33
$P(Y_{it}^{drink}=1\|Y_{i,t-1}^{drink}=0)$	31.73	33.22	41.10	33.12	35.04	43.75	31.73	31.92	36.88
$P(Y_{it}^{drink}=1\|Y_{i,t-1}^{drink}=1)$-$P(Y_{it}^{drink}=1\|Y_{i,t-1}^{drink}=0)$	56.01	54.55	23.57	57.01	54.32	13.34	56.01	56.02	22.45
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{mar}=1)$	66.99	66.48	35.63	69.06	68.21	34.92	66.99	66.30	36.85
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{mar}=0)$	9.09	9.26	11.12	10.45	10.72	12.76	9.09	8.80	10.41
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{mar}=1)$-$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{mar}=0)$	57.90	57.22	24.51	58.61	57.49	22.16	57.90	57.50	26.44
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{coc}=1)$	43.40	42.42	14.63	43.84	43.73	14.81	43.40	41.54	14.20
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{coc}=0)$	3.48	3.58	3.82	4.26	4.32	4.70	3.48	3.39	3.60
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{coc}=1)$-$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{coc}=0)$	39.92	38.84	10.81	39.58	39.41	10.10	39.92	38.15	10.60
Panel B: First Order Transitions from Softer to Harder Drugs (Stepping Stone)									
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{drink}=1)$	30.35	29.69	25.47	34.06	33.50	21.01	30.35	28.42	25.72
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{drink}=0)$	6.57	7.81	7.59	6.43	7.36	7.01	6.57	8.00	7.83
$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{drink}=1)$-$P(Y_{it}^{mar}=1\|Y_{i,t-1}^{drink}=0)$	23.79	21.87	17.89	27.63	26.14	13.99	23.79	20.43	17.89
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{mar}=1)$	18.95	17.86	13.78	20.51	20.65	14.24	18.95	16.44	12.15
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{mar}=0)$	2.14	2.46	2.41	2.52	2.59	2.48	2.14	2.58	2.50
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{mar}=1)$-$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{mar}=0)$	16.80	15.40	11.37	17.99	18.06	11.76	16.80	13.85	9.65
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{drink}=1)$	8.49	8.35	6.18	9.87	9.84	5.44	8.49	7.89	6.55
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{drink}=0)$	1.28	1.57	1.53	1.54	1.67	1.72	1.28	1.49	1.45
$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{drink}=1)$-$P(Y_{it}^{coc}=1\|Y_{i,t-1}^{drink}=0)$	7.21	6.78	4.66	8.33	8.17	3.72	7.21	6.39	5.10

Note: Model A, B, and C are described in section 3.
The column labeled Sim($\gamma_{kj}=0$) presents probabilities that were estimated "turning off" the relevant γ_{kj}, and holding all other parameters constant.

Table 5: Comparison of Predicted and Observed Drug-Use Participation Sequences

Sum	Transitions	Model A		Model B		Model C	
		Obs	Pred	Obs	Pred	Obs	Pred
Panel A: Alcohol Use Participation Sequences							
0		6.07	4.34	5.40	3.61	6.07	4.15
1	1	1.43	1.71	1.20	1.29	1.43	1.67
1	2	2.56	2.94	2.00	2.25	2.56	2.84
2,3	1	2.11	1.88	2.40	1.64	2.11	3.03
2,3	2	1.66	2.89	0.80	2.20	1.66	2.13
2,3	3	2.33	2.52	1.70	1.98	2.33	2.72
2,3	4,6	2.60	1.82	1.50	1.16	2.60	2.47
4,5	1	3.58	2.20	4.50	2.78	3.58	4.43
4,5	2	0.90	1.90	0.40	1.91	0.90	2.60
4,5	3	3.15	3.51	2.70	3.25	3.15	3.74
4,5	4,6	4.27	3.08	3.20	2.00	4.27	3.11
4,5	7,9	0.20	0.08	0.10	0.04	0.20	0.26
6,7	1	6.03	5.31	5.70	6.59	6.03	5.89
6,7	2	1.98	3.43	2.40	4.17	1.98	3.27
6,7	3	5.25	6.25	5.00	5.69	5.25	4.62
6,7	4,6	5.11	4.39	4.30	3.53	5.11	4.21
6,7	7,9	0.27	0.05	0.10	0.00	0.27	0.11
8,10	0	24.39	21.24	28.90	22.92	24.39	22.87
8,10	1	10.85	11.98	11.00	12.71	10.85	10.27
8,10	2	10.30	13.19	11.20	15.49	10.30	10.67
8,10	3	2.88	3.51	3.20	2.97	2.88	3.04
8,10	4,6	2.08	1.79	2.30	1.82	2.08	1.91
Total		100.00	100.00	100.00	100.00	100.00	100.00

Panel B: Marijuana Use Participation Sequences

0		45.54	44.38	41.90	38.30	45.54	46.93
1	1	3.86	4.16	4.00	4.80	3.86	3.62
1	2	10.10	11.06	9.10	11.30	10.10	10.36
2,3	1	2.55	2.22	2.40	2.70	2.55	2.15
2,3	2	4.46	4.63	3.70	5.90	4.46	4.45
2,3	3	3.33	3.51	2.90	3.30	3.33	3.61
2,3	4,6	4.80	3.46	4.70	2.90	4.80	4.30
4,5	1	1.45	1.07	1.20	1.30	1.45	1.21
4,5	2	1.41	2.14	1.10	2.00	1.41	2.05
4,5	3	2.56	3.13	3.60	3.30	2.56	2.32
4,5	4,6	3.70	4.07	5.10	3.00	3.70	3.35
4,5	7,9	0.27	0.11	0.40	0.10	0.27	0.19
6,7	1	1.70	1.14	2.80	1.60	1.70	1.46
6,7	2	0.98	1.64	1.10	2.70	0.98	1.42
6,7	3	2.53	2.62	2.90	3.40	2.53	2.19
6,7	4,6	2.25	2.44	2.90	3.30	2.25	2.10
6,7	7,9	0.00	0.04	0.00	0.00	0.00	0.07
8,10	0	2.72	2.29	3.20	1.70	2.72	2.04
8,10	1	2.66	2.18	2.90	3.00	2.66	2.48
8,10	2	1.68	2.31	1.70	3.40	1.68	2.17
8,10	3	0.94	0.99	1.40	1.40	0.94	0.96
8,10	4,6	0.51	0.41	1.00	0.60	0.51	0.57
Total		100.00	100.00	100.00	100.00	100.00	100.00

Panel C: Cocaine Use Participation Sequences

0		75.86	73.83	72.40	68.50	75.86	75.55
1	1	2.47	2.80	3.00	2.96	2.47	2.83
1	2	8.40	9.88	8.20	9.77	8.40	9.19
2,3	1	1.04	1.03	1.00	1.42	1.04	0.96
2,3	2	2.74	3.47	3.00	5.30	2.74	2.67
2,3	3	1.59	1.69	2.40	1.92	1.59	1.44
2,3	4,6	2.82	2.55	3.70	3.33	2.82	2.60
4,5		3.09	2.93	3.70	5.32	3.09	3.05
6,7		1.41	1.44	1.90	1.34	1.41	1.36
8,10		0.59	0.39	0.70	0.14	0.59	0.34
Total		100.00	100.00	100.00	100.00	100.00	100.00

Note: I divided the potential drug sequences into cells that can be described by their "sum" (number of periods in which the drug was used) and "transitions" (the number of times the sequences transitioned from 0 to 1 or from 1 to 0. Section 3 describes models A, B, and C. This table reports the observed and predicted share of the sample that corresponds to each cell for alcohol, marijuana, and cocaine.

Table 6: Comparison of Predicted and Observed Combinations of Drug Use by Year

		Model A		Model B		Model C	
		Obs	Pred	Obs	Pred	Obs	Pred
Bundle	year						
(0,0,0)	1998	49.10	47.02	45.80	44.10	49.10	44.95
(1,0,0)	1998	29.97	31.34	30.00	30.80	29.97	32.51
(0,1,0)	1998	1.62	2.51	2.10	2.60	1.62	5.24
(0,0,1)	1998	0.12	0.23	0.20	0.40	0.12	0.78
(1,1,0)	1998	14.35	14.13	16.40	17.40	14.35	12.33
(0,1,1)	1998	0.10	0.16	0.20	0.20	0.10	0.39
(1,0,1)	1998	0.80	0.94	1.00	1.30	0.80	2.07
(1,1,1)	1998	3.94	3.68	4.30	3.10	3.94	1.74
(0,0,0)	1999	44.54	40.31	40.40	40.60	44.54	41.39
(1,0,0)	1999	33.03	34.55	34.60	32.20	33.03	34.75
(0,1,0)	1999	1.55	4.35	1.10	2.80	1.55	4.52
(0,0,1)	1999	0.27	0.65	0.00	0.50	0.27	0.78
(1,1,0)	1999	14.84	14.55	17.40	17.10	14.84	13.42
(0,1,1)	1999	0.23	0.41	0.50	0.40	0.23	0.42
(1,0,1)	1999	0.80	1.64	0.80	1.10	0.80	2.44
(1,1,1)	1999	4.74	3.50	5.20	5.40	4.74	2.28
(0,0,0)	2000	39.17	36.22	36.90	36.30	39.17	37.03
(1,0,0)	2000	35.22	37.20	35.70	35.70	35.22	36.96
(0,1,0)	2000	1.64	4.07	1.30	2.10	1.64	4.16
(0,0,1)	2000	0.20	0.53	0.20	0.30	0.20	0.57
(1,1,0)	2000	17.27	15.78	18.00	17.40	17.27	15.68
(0,1,1)	2000	0.27	0.37	0.40	0.30	0.27	0.37
(1,0,1)	2000	0.76	1.68	1.00	1.00	0.76	1.86
(1,1,1)	2000	5.46	4.17	6.50	6.90	5.46	3.38
(0,0,0)	2001	34.32	31.52	31.80	30.80	34.32	32.92
(1,0,0)	2001	39.51	41.60	38.80	40.50	39.51	40.80
(0,1,0)	2001	1.74	3.21	0.60	1.50	1.74	3.47
(0,0,1)	2001	0.12	0.47	0.00	0.30	0.12	0.40
(1,1,0)	2001	17.42	16.72	19.60	18.30	17.42	16.40
(0,1,1)	2001	0.14	0.31	0.00	0.30	0.14	0.31
(1,0,1)	2001	0.80	1.82	0.90	1.10	0.80	1.88
(1,1,1)	2001	5.95	4.39	8.30	7.40	5.95	3.84
(0,0,0)	2002	29.80	27.56	27.90	26.40	29.80	28.89
(1,0,0)	2002	45.11	45.73	43.40	45.20	45.11	45.52
(0,1,0)	2002	1.06	2.60	1.00	1.10	1.06	2.83
(0,0,1)	2002	0.14	0.35	0.10	0.10	0.14	0.29
(1,1,0)	2002	17.62	17.25	19.40	18.50	17.62	16.57
(0,1,1)	2002	0.14	0.25	0.10	0.10	0.14	0.23
(1,0,1)	2002	1.06	1.78	1.40	1.10	1.06	1.72
(1,1,1)	2002	5.09	4.46	6.70	7.40	5.09	3.94

(0,0,0)	2003	27.62	24.94	23.10	22.10	27.62	26.23
(1,0,0)	2003	48.63	49.47	47.50	50.00	48.63	49.38
(0,1,0)	2003	1.12	2.08	0.60	0.80	1.12	2.30
(0,0,1)	2003	0.06	0.29	0.10	0.20	0.06	0.27
(1,1,0)	2003	16.48	17.05	20.80	19.00	16.48	16.19
(0,1,1)	2003	0.12	0.18	0.20	0.10	0.12	0.17
(1,0,1)	2003	1.14	1.82	1.50	1.00	1.14	1.78
(1,1,1)	2003	4.84	4.17	6.20	0.70	4.84	3.68
(0,0,0)	2004	25.80	23.02	21.50	20.10	25.80	24.15
(1,0,0)	2004	51.70	53.41	52.10	52.40	51.70	53.29
(0,1,0)	2004	0.86	1.49	0.50	0.50	0.86	1.76
(0,0,1)	2004	0.20	0.27	0.30	0.10	0.20	0.14
(1,1,0)	2004	15.84	16.33	18.80	19.30	15.84	15.36
(0,1,1)	2004	0.08	0.14	0.00	0.10	0.08	0.14
(1,0,1)	2004	1.10	1.70	1.50	1.10	1.10	1.63
(1,1,1)	2004	4.42	3.64	5.30	6.50	4.42	3.53
(0,0,0)	2005	22.28	21.87	19.20	19.30	22.28	22.71
(1,0,0)	2005	55.81	55.58	56.10	54.80	55.81	55.90
(0,1,0)	2005	0.61	1.27	0.40	0.50	0.61	1.41
(0,0,1)	2005	0.14	0.25	0.10	0.00	0.14	0.15
(1,1,0)	2005	15.58	16.01	18.50	18.50	15.58	15.06
(0,1,1)	2005	0.08	0.10	0.00	0.00	0.08	0.11
(1,0,1)	2005	0.94	1.78	0.70	1.10	0.94	1.60
(1,1,1)	2005	4.56	3.17	5.00	5.90	4.56	3.07
(0,0,0)	2006	21.73	21.06	17.10	18.30	21.73	21.64
(1,0,0)	2006	58.07	58.10	57.80	56.90	58.07	58.30
(0,1,0)	2006	0.70	1.16	0.50	0.30	0.70	1.23
(0,0,1)	2006	0.06	0.20	0.00	0.10	0.06	0.14
(1,1,0)	2006	14.62	14.92	19.10	17.80	14.62	14.29
(0,1,1)	2006	0.02	0.08	0.00	0.00	0.02	0.10
(1,0,1)	2006	1.06	1.64	1.00	1.10	1.06	1.69
(1,1,1)	2006	3.74	2.86	4.50	5.60	3.74	2.61
(0,0,0)	2007	21.97	20.50	16.00	17.10	21.97	20.90
(1,0,0)	2007	59.14	59.67	60.60	59.20	59.14	60.18
(0,1,0)	2007	0.70	0.96	0.40	0.30	0.70	1.19
(0,0,1)	2007	0.06	0.18	0.00	0.00	0.06	0.13
(1,1,0)	2007	14.06	14.51	17.30	17.70	14.06	13.56
(0,1,1)	2007	0.04	0.06	0.00	0.00	0.04	0.07
(1,0,1)	2007	1.02	1.66	1.50	1.00	1.02	1.62
(1,1,1)	2007	3.01	2.47	4.20	4.70	3.01	2.35

Note: The row corresponding to $(Y_{it}^{drink}, Y_{it}^{mar}, Y_{it}^{coc})$ indicates the actual and predicted share of the sample who consumed that particular bundle in year t. For instance, the second to last row (1,0,1) corresponds to the bundle where a respondent consumed alcohol, did not consume marijuana, and consumed cocaine in the same period.

Table 7: Sample Analogue of Generalized Residuals

	Model A: Logit M=7 (1st Order)			Model B: Probit M=3			Model C: Logit (2nd Order)		
Panel A: Mean, Variance, and Serial Correlation of Generalized Residuals									
	drink	mar	coc	drink	mar	coc	drink	mar	coc
$E[r_{it}^j(\alpha_i)]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
$E[r_{it}^j(\alpha_i)^2]$	0.99	0.97	0.98	1.00	0.95	1.05	1.00	0.97	1.01
	(0.01)	(0.02)	(0.05)	(0.04)	(0.04)	(0.14)	(.01)	(.02)	(.06)
$E[r_{it}^j(\alpha_i), r_{i,t-1}^j(\alpha_i)]$	-0.03	-0.02	-0.01	-0.04	-0.02	-0.02	0.03	0.03	0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-2}^j(\alpha_i)]$	0.08	0.05	0.04	0.08	0.05	0.11	-0.04	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.07)	(.01)	(0.00)	(.01)
$E[r_{it}^j(\alpha_i), r_{i,t-3}^j(\alpha_i)]$	0.05	0.03	0.02	0.08	0.05	0.05	0.06	0.02	0.02
	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)	(.01)	(0.00)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-4}^j(\alpha_i)]$	0.03	0.02	0.02	0.06	0.06	0.09	0.05	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(.01)	(.01)	(.01)
$E[r_{it}^j(\alpha_i), r_{i,t-5}^j(\alpha_i)]$	0.01	0.01	0.01	0.04	0.00	0.04	0.05	0.01	0.01
	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(.01)	(.01)	(0.00)
Panel B: Correlation of Generalized Residuals Across Drugs and Time									
	coc-drink	mar-drink	coc-mar	coc-drink	mar-drink	coc-mar	coc-drink	mar-drink	coc-mar
$E[r_{it}^j(\alpha_i), r_{i,t}^{j'}(\alpha_i)]$	0.04	0.11	0.10	0.07	0.18	0.17	0.05	0.13	0.12
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)
$E[r_{it}^j(\alpha_i), r_{i,t-1}^{j'}(\alpha_i)]$	0.00	-0.01	0.00	-0.02	-0.02	-0.01	0.00	0.02	0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-2}^{j'}(\alpha_i)]$	0.00	0.01	0.00	0.03	0.03	0.01	-0.01	0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-3}^{j'}(\alpha_i)]$	0.01	0.00	0.00	0.00	0.04	0.01	0.01	0.00	0.00
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-4}^{j'}(\alpha_i)]$	0.00	-0.01	-0.01	0.00	0.04	0.03	0.00	-0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$E[r_{it}^j(\alpha_i), r_{i,t-5}^{j'}(\alpha_i)]$	0.00	-0.01	-0.01	0.03	0.02	0.00	0.00	-0.01	0.00
	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)

Note: See section 4D for a description of generalized residuals diagnostics. Panel A corresponds to the mean, variance and 1st-5th order autocorrelation of sample-analogue generalized residuals for alcohol, marijuana, and cocaine. Panel B corresponds to the sample-analogue of generalized residuals across drugs.

Table 8:Parameter Estimates for Models with Heterogeneous Stepping-Stone

	Model D: By Age			Model E: By Gender			Model F: By RE of Y			Model G: by RE of X			Model H: by Intensity			Model I: by Accumulation		
	Alcohol	Mar	Coc	Alcohol	Mar	Coc	Alcohol	Mar	Coc	Alcohol	Mar	Coc	Alcohol	Mar	Coc	Alcohol	Mar	Coc
Ldrink cons	-0.50	2.23	1.79	1.80	0.23	0.17	1.68	0.39	0.52	1.68	0.34	0.49				1.756	0.406	0.366
	(0.20)	(0.26)	(0.44)	(0.04)	(0.07)	(0.12)	(0.08)	(0.09)	(0.13)	(0.08)	(0.09)	(0.13)				(0.044)	(0.059)	(0.100)
Lmarcons	0.93	-0.32	1.32	0.02	1.77	0.61	0.45	1.83	0.48	0.48	1.84	0.48				0.179	1.591	0.463
	(0.31)	(0.25)	(0.37)	(0.07)	(0.05)	(0.08)	(0.10)	(0.06)	(0.07)	(0.10)	(0.06)	(0.07)				(0.071)	(0.058)	(0.084)
Lcoc cons	1.79	1.02	-0.25	0.11	0.22	1.80	0.17	0.16	1.83	0.18	0.17	1.83				-0.191	0.027	1.604
	(0.56)	(0.41)	(0.40)	(0.12)	(0.08)	(0.08)	(0.14)	(0.07)	(0.06)	(0.14)	(0.07)	(0.06)				(0.119)	(0.084)	(0.078)
Ldrink by (*)	0.12	-0.11	-0.09	0.17	0.06	0.29	-0.17	0.16	0.33	-0.17	0.22	0.52				0.025	-0.029	-0.011
	(0.01)	(0.01)	(0.02)	(0.06)	(0.09)	(0.16)	(0.05)	(0.07)	(0.14)	(0.05)	(0.12)	(0.22)				(0.007)	(0.007)	(0.010)
Lmar by (*)	-0.05	0.11	-0.04	-0.07	0.10	-0.14	0.68	0.01	-0.22	0.38	0.01	-0.18				-0.058	0.058	0.021
	(0.02)	(0.01)	(0.02)	(0.09)	(0.07)	(0.11)	(0.17)	(0.05)	(0.11)	(0.07)	(0.05)	(0.09)				(0.014)	(0.010)	(0.013)
Lcoc by demog(*)	-0.10	-0.04	0.10	-0.37	-0.17	0.00	0.69	0.10	0.27	0.43	0.11	0.26				0.046	0.034	0.058
	(0.03)	(0.02)	(0.02)	(0.16)	(0.11)	(0.11)	(0.20)	(0.09)	(0.15)	(0.13)	(0.10)	(0.15)				(0.033)	(0.019)	(0.016)
Ldrink (low dose)													1.62	0.26	0.29			
													(0.03)	(0.05)	(0.09)			
Lmar (low dose)													-0.02	1.43	0.42			
													(0.04)	(0.04)	(0.07)			
Lcoc (low dose)													0.06	0.14	1.54			
													(0.06)	(0.06)	(0.07)			
Ldrink (high dose)													2.97	0.31	0.63			
													(0.04)	(0.06)	(0.10)			
Lmar (high dose)													-0.06	3.14	0.77			
													(0.05)	(0.05)	(0.07)			
Lcoc (high dose)													0.14	-0.02	2.11			
													(0.07)	(0.07)	(0.07)			
c1													-0.51	1.13	2.78			
													(0.08)	(0.11)	(0.17)			
c2													2.85	3.14	3.91			
													(0.08)	(0.11)	(0.17)			
LL		-48650			-48792			-48766			-48765			-69395			-48652	
Nparams		68			68			68			68			74			68	
N		5108			5108			5108			5108			4847			5100	

The stepping-stone effect from drug k to drug j is $\gamma_{kj} = \gamma_{kj}^{cons} + \gamma_{kj}^*(*)$, where (*) corresponds to age at time t for model D, gender for model E, α_j for Model F, α_k for model G, and the number of years that have elapsed since first use of drug j as a measure of addiction capital accumulation for Model I. The stepping-stone effect for model H varies by intensity of consumption in the last period, where Ldrink(low dose) and Ldrink(high dose) is the stepping-stone effect from low and high levels of alcohol in the past period, respectively.

Table 9:Parameter Estimates for Models with Time and Drug Specific Covariates

	Treatment Admissions			Prices (Low Mar, Low Crack Cocaine)			Prices (High Mar, High Crack Cocaine)		
	Alcohol	**Marijuana**	**Cocaine**	**Alcohol**	**Marijuana**	**Cocaine**	**Alcohol**	**Marijuana**	**Cocaine**
Lag alcohol	1.885	0.316	0.413	1.885	0.316	0.414	1.885	0.317	0.414
	(0.035)	(0.058)	(0.100)	(0.035)	(0.058)	(0.100)	(0.035)	(0.058)	(0.100)
Lag mar	-0.006	1.838	0.489	-0.006	1.838	0.489	-0.006	1.839	0.489
	(0.059)	(0.043)	(0.063)	(0.059)	(0.043)	(0.063)	(0.059)	(0.043)	(0.063)
Lag coc	-0.039	0.090	1.799	-0.038	0.090	1.799	-0.038	0.090	1.799
	(0.093)	(0.061)	(0.059)	(0.093)	(0.061)	(0.059)	(0.093)	(0.061)	(0.059)
LL		42256.7			42257.4			42258.1	

Note: I estimate these parameters in a model with drug and time specific variables. For the first three columns, I use state and year level deviations from national trends in treatment admissions for substance abuse for a particular drug (for the state in which respondent i resided in year t) as exclusion restrictions for alcohol, marijuana and cocaine. For columns 4 to 9 I interact this time and drug specific measure of substance abuse with time-varying prices for each drug as the exclusion restriction. For columns 4, 5 and 6 I use price per gram of marijuana and crack cocaine when purchased in low volumes. For columns 7, 8 and 9, I use price per gram of marijuana and crack cocaine when purchased in high volumes. The latent utility for alcohol in every specification also included a time-varying indicator for whether the respondent was at least 21 years old at the time of the interview.

Table 10:Parameter Estimates for Models that Allow Complementarity

	Alc+Mar		Alc+Coc		Mar+Coc	
	Alcohol	Marijuana	Alcohol	Cocaine	Marijuana	Cocaine
Panel A: Without Prices						
Lag alcohol	1.802	0.202	1.804	0.497	-	-
	(0.033)	(0.050)	(0.032)	(0.093)		
Lag mar	-0.109	1.778	-	-	1.598	0.057
	(0.051)	(0.039)			(0.042)	(0.064)
Lag cocaine	-	-	0.045	1.509	-0.219	1.917
			(0.093)	(0.066)	(0.067)	(0.057)
T	1.761		1.522		1.571	
	(0.058)		(0.116)		(0.070)	
LL	39965.6		31179.7		26111.9	
Panel B: With Prices						
Lag alcohol	1.813	0.235	1.820	0.545	-	-
	(0.035)	(0.054)	(0.034)	(0.100)		
Lag mar	-0.119	1.776	-	-	1.607	0.007
	(0.055)	(0.041)			(0.045)	(0.067)
Lag cocaine	-	-	0.049	1.495	-0.271	1.929
			(0.100)	(0.068)	(0.070)	(0.061)
T	1.767		1.419		1.546	
	(0.062)		(0.122)		(0.074)	
LL	34440.8		26868.1		22862.9	

Note: I extend the model of complementarity first developed by Gentzkow (2007) to allow for state dependence within and between drugs for all possible pairs. The two drugs are complements if T>0 and they are substitutes if T<0. Panel B includes time and drug specific covariates.

Table A1: Summary Statistics Among Subsamples

	Full Sample		Sample not lost due to attrition 1997-2007			
			Answered Drug Questions for all 10 waves			
			All	Alcohol	Mar	Coc
	(1)	(2)	(3)	(4)	(5)	(6)
Age(1997)	14.35	14.23	14.23	14.24	14.23	14.23
	(1.49)	(1.47)	(1.47)	(1.47)	(1.47)	(1.47)
Male	0.51	0.48	0.47	0.47	0.47	0.47
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Father in household	0.72	0.72	0.73	0.73	0.73	0.73
	(0.45)	(0.45)	(0.44)	(0.45)	(0.44)	(0.44)
Age First Drink	14.68	14.95	15.03	14.99	15.01	14.99
	(3.42)	(3.40)	(3.41)	(3.41)	(3.41)	(3.42)
Age First Marijuana	15.69	16.08	16.17	16.15	16.16	16.15
	(3.02)	(3.03)	(3.01)	(3.01)	(3.02)	(3.01)
Age First Cocaine	17.07	17.48	17.58	17.56	17.56	17.59
	(3.32)	(3.42)	(3.32)	(3.34)	(3.34)	(3.32)
Percentage of respondents with non-missing answers for drug-related questions(a)						
P(non-missing alcohol)	59.21	94.59	100.00	100.00	98.77	98.14
P(non-missing marijuana)	58.07	92.78	100.00	96.88	100.00	97.29
P(non-missing cocaine)	59.13	94.47	100.00	98.01	99.06	100.00
Percentage of respondents who reported having consumed drugs at least once (b)						
P(ever alcohol)	92.39	94.59	94.58	94.75	94.58	94.65
P(ever marijuana)	57.66	58.69	57.34	58.32	57.41	57.79
P(ever cocaine)	23.12	25.29	24.63	25.02	24.80	24.60
N	8984	5623	5112	5319	5217	5312

Note: Standard errors are in parenthesis.
Column (1) corresponds to the full sample. Column (2) corresponds to the subsample that was not lost due to attrition, whether or not they had non-missing answers to the drug-related questions. Columns (4)-(5)-(6) are restricted to respondents who are not lost due to attrition and had non-missing answers to alcohol, marijuana, and cocaine, respectively. Column (3) corresponds to the intersection of (4)-(5)-(6).

Table A2: Comparison Between Estimated With and Without Classification Error

	No Measurement Error			Measurement Error		
	Alcohol	Marijuana	Cocaine	Alcohol	Marijuana	Cocaine
Lag alcohol	1.74	0.28	0.33	1.70	0.23	0.30
	(0.03)	(0.05)	(0.09)	(0.03)	(0.05)	(0.09)
Lag marijuana	0.07	1.66	0.30	-0.07	1.75	0.30
	(0.05)	(0.04)	(0.06)	(0.05)	(0.04)	(0.06)
Lag cocaine	-0.08	-0.04	1.58	-0.11	-0.05	1.24
	(0.09)	(0.06)	(0.06)	(0.09)	(0.06)	(0.06)

Note: The first three columns correspond to the stepping-stone parameters estimated by Model A $\{\theta_{m=7}\}$ under the assumption of no misclassification error. The last three columns correspond to the stepping-stone parameters estimated under the assumption that $\{\theta_{m=7}\}$ are true, but every year, 20% of cocaine users, chosen at random, report not having used cocaine in the last year.

Table A3: Comparing the NLSY97 with Other Sources of Data Among Young Adults (18-25) in 2002

	$NLSY97^{(a)}$	$NSDUH^{(b)}$	$MTF^{(C)}$
Min Age in 2002	18	18	19
Max Age in 2002	23	25	24
Lifetime Drug Use			
Lifetime Alcohol	86.23	86.70	88.40
Lifetime Marijuana	52.52	53.80	56.10
Lifetime Cocaine (*)	18.67	15.40	12.90
Past Year Drug Use			
Alcohol	67.65	77.90	83.90
Marijuana	24.51	29.80	34.20
Cocaine (*)	6.03	6.70	6.50
Past Month Drug Use			
Alcohol	56.98	60.50	67.70
Marijuana	18.57	17.30	19.80
Cocaine (*)	-	2.00	2.50

Source:
(a)National Longitudinal Study of Youth 1997 (NLSY97)
(b)National Survey of Drug Use and Health (NSDUH)
(c)Monitoring the Future (MTF)
Data for the NSDUH and MTF correspond to year 2002 in Table 8.2 from Substance Abuse and Mental Health Services Administration, *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings*, NSDUH Series H-41, HHS Publication No. (SMA) 11-4658. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2011.
(*)Cocaine in the NLSY97 is grouped with other hard drugs.

Table A4: Parameter Estimates for the Benchmark Model (Model A) with more Covariates

Panel A: Time-Invariant Covariates

	Alcohol	Marijuana	Cocaine		Alcohol	Marijuana	Cocaine		Alcohol	Marijuana	Cocaine
Lag alcohol	1.878	0.258	0.316		1.868	0.280	0.253		1.922	0.293	0.247
	(0.032)	(0.052)	(0.089)		(0.034)	(0.052)	(0.092)		(0.033)	(0.050)	(0.090)
Lag mar	-0.017	1.826	0.543		-0.043	1.788	0.512		-0.055	1.773	0.520
	(0.053)	(0.040)	(0.059)		(0.054)	(0.040)	(0.061)		(0.052)	(0.039)	(0.061)
Lag Coc	-0.085	0.136	1.800		-0.149	0.119	1.739		-0.164	0.129	1.745
	(0.084)	(0.057)	(0.055)		(0.087)	(0.060)	(0.057)		(0.085)	(0.059)	(0.057)
black		N				Y				Y	
hispanic		N				Y				Y	
Urban in R1		N				N				Y	

Panel B: Time-Varying Covariates

	Alcohol	Marijuana	Cocaine		Alcohol	Marijuana	Cocaine		Alcohol	Marijuana	Cocaine
Lag alcohol	1.886	0.208	0.338		1.896	0.239	0.335		1.881	0.265	0.350
	(0.034)	(0.055)	(0.093)		(0.034)	(0.054)	(0.093)		(0.034)	(0.056)	(0.095)
Lag mar	-0.063	1.841	0.494		-0.073	1.830	0.505		-0.036	1.839	0.506
	(0.055)	(0.042)	(0.061)		(0.055)	(0.041)	(0.062)		(0.057)	(0.042)	(0.062)
Lag Coc	-0.073	0.106	1.800		-0.094	0.105	1.792		-0.066	0.113	1.830
	(0.089)	(0.060)	(0.057)		(0.088)	(0.060)	(0.057)		(0.089)	(0.060)	(0.057)
enr. 2yr coll		Y				N				N	
enr4yrcoll		Y				N				N	
curr emp		Y				N				N	
unemp rate		N				N				Y	
region in t		N				Y				N	

Note: These parameters are estimated using a multivariate logit with three mass points (includes all features of the benchmark model but it only uses three mass points). Panel A corresponds to a model with time-invariant covariates. The first three columns have the same specification as the benchmark model. Columns 4, 5 and 6 include an indicator for whether the respondent is black or Hispanic. Finally, the last three columns also include an indicator for whether the respondent lived in an urban area in the first wave. Panel B corresponds to a model with time-varying covariates. The first three columns include indicators for whether the individual is enrolled in a 2-year or 4-year college and whether the individual is employed. Columns 4, 5, and 6 include a time-varying indicator of region, but most individuals do not change region of residence during the survey. The last three columns control for the year specific unemployment rate that corresponds to the individual state of residence at the time of interview.

Table A5: Actual and Conterfactual Share of Respondents Who Consume Each Drug by Age Group

	13-14	15-16	17-18	19-20	21-22	23-24	25+
Panel A: Observed							
Alcohol	43.23	59.88	72.47	80.19	83.46	81.58	81.12
Mar	16.29	28.16	34.04	32.30	27.64	22.64	18.07
Coc	4.70	8.13	11.15	10.06	8.73	6.77	4.94
Panel B: Simulated							
Alcohol	42.97	58.94	71.58	79.88	82.24	80.70	81.54
	(1.02)	(0.64)	(0.52)	(0.41)	(0.40)	(0.46)	(0.75)
Mar	17.50	27.61	32.79	30.51	27.35	21.92	19.23
	(0.75)	(0.54)	(0.59)	(0.48)	(0.48)	(0.50)	(0.76)
Coc	4.56	8.12	10.63	9.83	8.32	6.22	4.92
	(0.47)	(0.39)	(0.35)	(0.36)	(0.35)	(0.33)	(0.48)
Panel C: Conterfactual Scenario where Alcohol and Marijuana are Removed Until Age 18							
Alcohol	0.00	0.00	0.00	72.67	81.27	80.50	81.49
	(0.00)	(0.00)	(0.00)	(0.48)	(0.42)	(0.46)	(0.75)
Mar	0.00	0.00	0.00	23.62	26.22	21.69	19.17
	(0.00)	(0.00)	(0.00)	(0.48)	(0.50)	(0.51)	(0.76)
Coc	3.82	5.81	6.91	6.97	7.63	6.07	4.89
	(0.44)	(0.32)	(0.32)	(0.34)	(0.36)	(0.33)	(0.47)
Panel D: Conterfactual Scenario Where Alcohol an dMarijuana are Removed Until Age 20							
Alcohol	0.00	0.00	0.00	0.00	71.97	78.72	81.08
	(0.00)	(0.00)	(0.00)	(0.00)	(0.53)	(0.50)	(0.76)
Mar	0.00	0.00	0.00	0.00	21.24	20.85	18.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.55)	(0.51)	(0.78)
Coc	3.82	5.81	6.91	6.27	6.41	5.75	4.82
	(0.44)	(0.32)	(0.32)	(0.31)	(0.33)	(0.32)	(0.47)
Panel E: Conterfactual Scenario Where Alcohol an dMarijuana are Removed Until Age 22							
Alcohol	0.00	0.00	0.00	0.00	0.00	65.95	78.69
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.62)	(0.77)
Mar	0.00	0.00	0.00	0.00	0.00	17.12	18.52
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.46)	(0.74)
Coc	3.82	5.81	6.91	6.27	6.16	5.45	4.73
	(0.44)	(0.32)	(0.32)	(0.31)	(0.33)	(0.32)	(0.44)

Note: This table presents the simulated share of cocaine users by age group and the corresponding standard errors if alcohol and marijuana were artificially removed until age 18 (Panel C), age 20 (Panel D) and age 22 (Panel E). These estimates correspond to Figure 4. I calculate the standard errors as follows. First, I simulate drug outcomes using the parameters estimated in the model and the demographics of my entire sample. Second, I use those simulated drug outcomes to obtain an estimate for the share of individuals in the sample that my model predicts will be using drugs by each age category (μ_{jatr} is the predicted share of individuals that consume drug j in age category a, and simulation repetition r for year t). Third, I repeat this process 100 times. Then, I compute a mean and standard deviation for $(\mu_{jat1}, \mu_{jat2}, \ldots \ldots, \mu_{jat100})$ and report those as my standard errors.

Table A6: Actual and Conterfactual Share of Respondents who consumed Cocaine by Intensity of Use

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Panel A: Observed Data										
Coc None	95.23	94.18	93.54	93.05	93.71	94.08	94.41	94.66	95.42	96.12
Coc Low	1.57	3.59	4.00	4.00	3.88	3.49	3.05	3.07	2.43	2.23
Coc High	3.20	2.23	2.46	2.95	2.41	2.43	2.54	2.27	2.15	1.65
Cocaine Total	4.77	5.82	6.46	6.95	6.29	5.92	5.59	5.34	4.58	3.88
Panel B: Simulated Data										
Cocaine Total	4.78	5.68	6.15	6.34	6.26	5.92	5.34	4.82	4.29	3.85
	(0.26)	(0.30)	(0.34)	(0.30)	(0.34)	(0.31)	(0.32)	(0.29)	(0.27)	(0.25)
Panel C:Conterfactual Scenario where Alcohol Consumption in High Amounts is Replaced with Low Amounts										
Cocaine Total	4.78	5.45	5.73	5.75	5.54	5.10	4.48	3.96	3.49	3.10
	(0.26)	(0.29)	(0.32)	(0.32)	(0.32)	(0.30)	(0.29)	(0.27)	(0.23)	(0.24)
Panel D:Conterfactual Scenario where Marijuana Consumption in High Amounts is Replaced with Low Amounts										
Cocaine Total	4.78	5.51	5.73	5.69	5.46	5.06	4.49	4.05	3.62	3.26
	(0.26)	(0.30)	(0.31)	(0.31)	(0.32)	(0.30)	(0.31)	(0.29)	(0.25)	(0.24)
Panel E:Conterfactual Scenario where Alcohol and Marijuana Consumption in High Amounts is Replaced with Low Amounts										
Cocaine Total	4.78	5.28	5.32	5.13	4.81	4.36	3.76	3.34	2.97	2.64
	(0.26)	(0.29)	(0.31)	(0.31)	(0.29)	(0.29)	(0.28)	(0.26)	(0.22)	(0.21)
Panel F: Conterfactual Scenario Where High Levels of Marijuana are Replaced with No consumption at all										
Cocaine Total	4.78	5.34	5.39	5.24	5.00	4.63	4.11	3.73	3.36	3.04
	(0.26)	(0.28)	(0.31)	(0.30)	(0.31)	(0.28)	(0.29)	(0.27)	(0.24)	(0.23)
Panel G: Conterfactual Scenario where Marijuana (Low and High Amounts) is Removed Completely										
Cocaine Total	4.78	4.79	4.66	4.44	4.24	3.92	3.49	3.17	2.90	2.62
	(0.26)	(0.27)	(0.30)	(0.30)	(0.29)	(0.28)	(0.26)	(0.25)	(0.22)	(0.20)

Note: This table presents the simulated share of cocaine users by year and the corresponding standard errors under the counterfactual scenario where high levels of alcohol were replaced with low levels (Panel C), high levels of marijuana were replaced with low levels (Panel D), both high levels of alcohol and marijuana were replaced with low levels (Panel E), high levels of marijuana are replace with no marijuana use (Panel F) and finally a scenario where marijuana is removed completely (Panel G). These estimates correspond to Figure 5 and 6. I calculate the standard errors as follows. First, I simulate drug outcomes using the parameters estimated in the model and the demographics of my entire sample. Second, I use those simulated drug outcomes to obtain an estimate for the share of individuals in the sample that my model predicts will be using drugs by year ($\mu_{j,t,r}$ is the predicted share of individuals that consume drug j in year t, and simulation repetition r). Third, I repeat this process 100 times. Then, I compute a mean and standard deviation for ($\mu_{j,t,1}, \mu_{j,t,2}, \ldots \ldots, \mu_{j,t,100}$) and report those as my standard errors.