Title
Essays on Drug Use and Crime

Author
Deza, Monica

Publication Date
2012

Peer reviewed|Thesis/dissertation

Essays on Drug Use and Crime

By

Monica Deza

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Chair

Professor Patrick Kline

Professor Enrico Moretti

Professor Justin McCrary

Spring 2012

Essays on Drug Use and Crime

By

Monica Deza

Abstract

Essays on Drug Use and Crime

By

Monica Deza

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Card, Chair

This dissertation consists of three studies which analyze different aspects of risky behaviors and criminal participation.

A longstanding question is whether alcohol and marijuana use by teenagers exerts a "stepping stone" effect, increasing the chances that they will use harder drugs in the future. Empirically, teenagers who use alcohol or marijuana in one period are more likely to use cocaine in the future. This pattern can be explained in one of two ways: by a causal effect of soft drug consumption on future consumption of hard drugs (i.e., a true stepping-stone effect) or by unobserved characteristics that make people more likely to use soft drugs at a relatively young age, and hard drugs at a later age (i.e., correlated unobserved heterogeneity). Distinguishing between these alternatives is highly policy relevant because, to the extent that there is a true stepping stone effect, policies that reduce the use of soft drugs by young people will have lasting impact on the use of hard drugs by adults. In Chapter 1, I use data from the National Longitudinal Study of Youth 1997 (NLSY97) to estimate a dynamic discrete choice model of teenager's use of alcohol, marijuana and cocaine 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. In contrast, I find little evidence of a stepping stone effect from cocaine to alcohol or marijuana. Simulations show that restricting alcohol and marijuana use at young age has a modest impact on reducing later cocaine use.

Chapter 2 examines the role of an increase in alcohol consumption on drug initiation, hard drug consumption, and criminal participation. Using a regression discontinuity research design, I exploit the discontinuous increase in alcohol consumption at age 21 provided by the minimum legal drinking age. Using a survey of respondents during the year after they turned 21, I found that the probability of cocaine initiation decreased by 1.5 percentage points and the share of respondents who consumed cocaine in the last year decreased by 2 percentage points. Self-reported criminal participation, such as drug dealing, property destruction and attacking an individual, remained unchanged at age 21, with the exception of stealing, which decreased by 3 percentage points. These estimates are robust to a variety of specifications.

Between 1993 and 1995, a number of states implemented "Three Strikes and You're Out" laws that enhance the sentence length for repeat violent offenders. Chapter 3 develops a simple dynamic model that suggests that these laws will lead to an increase in the number of criminal cases that go to trial, rather than being settled with a plea bargain, since the threat of higher future sentences increases the cost of a being convicted for a strikeable offense. I use data from the 1990-2006 State Court Processing Statistics database and a state-by-year difference-in-differences research design to compare the change in the likelihood of plea bargaining by violent offenders after the passage of Three Strikes laws, relative to the trend among non-violent offenders. I also separately examine effects for offenders with at least one prior violent conviction, and compare the effects of the Three Strike law in California, which imposes extra sentencing for *any* third felony (violent or not), versus the eleven other states with Three Strikes laws. The results show that the introduction of Three Strikes laws significantly reduce the number of criminal cases that are settle with a plea bargain, imposing a potentially costly burden on the legal system.

Acknowledgements

I especially wish to thank David Card for being the best advisor I could wish for. I want to thank him for his dedication, patience, and encouragement. I am grateful to have had him as an undergraduate thesis advisor, dissertation chair, teacher, and role model.

I would also like to thank Steve Raphael, Justin McCrary, Enrico Moretti, Ronald Lee, Kenneth Train, and especially Patrick Kline for their continuous advice during these projects, for their support, and for their passion for teaching, which was reflected in the classes I took with them.

This research was supported by the UC Berkeley Integrated Graduate Education Research and Training in Politics, Economics, Psychology and Public Policy (IGERT), and by the 2011 ASHE Summer Dissertation Fellowship (SDF) Award.

This acknowledgements section would not be complete without thanking my friends and future colleagues from previous, current and future cohorts. I would like to thank Avraham Ebenstein, Kehinde Ajayi, Kevin Stange, Paulina Oliva, Jeffrey Greenbaum, and Raymundo Campos for their advice on how to survive graduate school. I would like to thank my cohort of labor economists Valentina Paredes, Edson Severnini, and Issi Romem for going through this journey with me, listening to my research and presentation several times, and providing very valuable advice at each stage of the project. I feel lucky to have them as colleagues. I would also like to thank colleagues in the younger cohort for their friendship and encouragement, especially Moises Yi, Michel Serafinelli, Marion Aouad, and Aaron Chalfin. I would also like to thank my students, especially Adrienne Sabety and Joseph Root, for showing me the rewarding side of teaching motivated and bright college students. I would also like to thank Cyndi Berck for excellent editing and proofreading services.

Finally, I would like to thank my father Franz Deza, my ‘mother’ Maty, and my best friend in the entire world, Juan, for his continuous support in this journey, from providing wake up calls to teaching me Beamer during job market season.

Contents

1. Is There a Stepping Stone Effect in Drug Use? Separating State Dependence from Unobserved Heterogeneity Within and Between Illicit Drugs

1.1 Introduction

Heated debates have arisen as states such as California 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) and many other outcomes. These models are also used in marketing to separate tastes from 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 (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. Throughout, I use mass-point mixing models to non-parametrically account for time-invariant multidimensional unobserved heterogeneity (Heckman and Singer, 1984).[1] My

[1] 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).

models also include flexible controls for the initial conditions problem caused by the fact that some individuals in the NLSY97 are first interviewed after they have already initiated soft (or even hard) drug use (Heckman 1981b; Wooldridge 2005).[2]

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 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. 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. Finally, 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.

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

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 sample-analogue of generalized residuals that allow 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, and presents counterfactual experiments. Finally, section 6 summarizes and concludes.

1.2. Data

I use data from the National Longitudinal Survey of Youth 1997 (NLSY97). This 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). [3] 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.[4] Measures of alcohol and marijuana use are available starting in 1997, while measures of cocaine use are available from 1998 onwards.[5]

Table 1 presents an overview of the characteristics of the sample. Column 1 shows data for the entire sample, including data on whether an individual has ever used the three main drugs that I focus on in this paper, and their age at first use. Column 2 summarizes my main analysis sample, which is restricted to people who were interviewed in every year from 1997 to 2007, and who also reported past-year drug consumption in all years from 1997 to 2007. 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.[6] Second, alcohol use is nearly universal (95% have ever used

[3] 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.

[4] 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.

[5] 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 since the last interview at least once and a 0 otherwise

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

alcohol) while marijuana is highly prevalent (57%) and cocaine 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.[7] Consistent with these patterns, the data on order of the first use of three substances in Panel B of Table 1 shows that, 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.

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[8]. 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. The key econometric goal of this paper is to explain how much of these patterns exhibited in Figure 1 and 2 arise because of true state dependence within and between drugs (stepping-stone effects), instead of correlated preferences.

1.3. 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).

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

[8] 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.

1.3.1. 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[9]. 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 age trend $(\delta_{0j} + \delta_{1j} t)$, on a set of observed characteristics (*X*) of the individual, [10] 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}:

$$U_{ijt} = \delta_{0j} + \delta_{1j}(t - t_0) + X_{it}\beta_j + \sum_{k=1}^{J} \gamma_{kj} Y_{i,k,t-1} + \alpha_{ij} + \varepsilon_{ijt} \tag{1.1}$$

I assume that the utility of 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). [11] I assume that the initial choices are selected according to the utility:

$$U_{ij0} = \omega_{0j} + X_{i0}\,\lambda_{0j} + \alpha_{ij0} + \varepsilon_{ijt}^{initial} \tag{1.2}$$

where again the utility of not consuming drug *j* in period *0* is set to 0. In equation (1) the state dependence and stepping stone effects are represented by the parameters γ_{kj}: for k=j these are the effects of lagged consumption of drug k on taste for that drug today. 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. 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). [12] Finally, ε_{ijt} and $\varepsilon_{ijt}^{initial}$ represent transitory taste

[9] 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 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.

[10] The vector X_{it} represents observable characteristics of consumer i in time t, such as age, sex, and whether the individual comes from a single-family household.

[11] 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).

[12] 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.

shocks which in the logistic model are assumed to be drawn independently from an Extreme Value Type 1 distribution.

Specification and Estimation

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, 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,[13]

$$P\left(Y_{ijt}=1\middle|\bar{Y}_{i,t-1},\alpha_{ij}\right)=\frac{\exp(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij})}{1+\exp(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij})} \tag{1.3a}$$

$$P\left(Y_{ijt}=0\middle|\bar{Y}_{i,t-1},\alpha_{ij}\right)=\frac{1}{1+\exp(\delta_{0,j}+\delta_{1j}(t-t_0)+X_{it}\beta_j+\sum_{k=1}^{J}\gamma_{kj}Y_{i,k,t-1}+\alpha_{ij})} \tag{1.3b}$$

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

$$\mathrm{P}\left(\mathrm{Y_{ij0}}=1\middle|\alpha_{\mathrm{ij0}}\right)=\frac{exp(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0})}{1+exp(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0})} \tag{1.4a}$$

$$P\left(Y_{ij0}=0\middle|\alpha_{ij0}\right)=\frac{1}{1+\exp(\omega_{0j}+X_{i0}\lambda_{0j}+\alpha_{ij0})} \tag{1.4b}$$

I compute the likelihood of a sequence of indicators of drug j consumption $L_i^j(Y_{ij0,\dots,}Y_{ijT})$ by taking the weighted average of type-specific likelihood contributions, using the unconditional probabilities, π_m, as weights.

$$L_i^j\left(Y_{ij0},..Y_{ijT}\right)=\sum_{m=1}^{M}\pi_m\{P(Y_{ij0}|\alpha_{mj0})\prod_{t=1999}^{T=2007}P(Y_{ijt}|\bar{Y}_{i,t-1},\alpha_{mj})\} \tag{1.5}$$

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

$$L_i=L_i^{alcohol}L_i^{marijuana}L_i^{cocaine} \tag{1.6}$$

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 ($\alpha_2^j,\dots,\alpha_7^j,\alpha_{1,0}^j,\dots,\alpha_{7,0}^j$)[14], 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 $(\phi_2,\dots,\phi_7)$.[15]

---

[13] $\bar{Y}_{i,t-1}=(Y_{i,t-1}^{drink},Y_{i,t-1}^{mar},Y_{i,t-1}^{coc})$, 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.

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

[15] 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

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

1.3.2. Model B: Trivariate Probit Model 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.[16] 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,

$$\left(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right) \sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & p_{12} & p_{13}\\ p_{12} & 1 & p_{23}\\ p_{13} & p_{23} & 1\end{pmatrix}\right) \tag{1.7}$$

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, $L_i(\bar{Y}_{i0,\ldots,}\bar{Y}_{iT})$, where

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

$$L_i(\bar{Y}_{i0}, \ldots . \bar{Y}_{iT}) = \sum_{m=1}^{M} \pi_m \{P(\bar{Y}_{i0} | \alpha_{m,0}) \prod_{t=1999}^{T=2007} P(\bar{Y}_{it} | \bar{Y}_{i,t-1}, \alpha_m)\} \tag{1.8}$$

The probabilities of a particular consumption bundle $\overline{Y_{it}}$, $P(\overline{\mathrm{Y}_{\mathrm{it}}} \mid \alpha_m)$ 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.[17] In addition to the parameters estimated under the logistic distribution, the multivariate probit model also estimates the correlation coefficients p_{12}, p_{13}, and p_{23}.

1.3.3. Model C: Trivariate Logit Model 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-

[16] 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. Since 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.

[17] For instance,

$$\begin{aligned} P\left(\bar{Y}_{it} = (1,1,0) \middle| \bar{Y}_{i,t-1}, \alpha_m\right) &= P(-\varepsilon_{it}^{drink} < V_{it}^{drink}(\alpha_m), -\varepsilon_{it}^{mar} < V_{it}^{mar}(\alpha_m), \varepsilon_{it}^{coc} < -V_{it}^{coc}(\alpha_m)) \\ &= \phi(V_{it}^{drink}(\alpha_m), V_{it}^{mar}(\alpha_m), -V_{it}^{coc}(\alpha_m); p_{12}, -p_{13}, -p_{23}) \end{aligned}$$

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

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 since section 6 indicates that correlation 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. Now 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.[18] Each initial conditions equation has a separate location parameter[19]. The latent utilities for each drug j are as follows:

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

(1.9)

$$U_{ij0}(0,0) = \varepsilon_{ij0}^{(0,0)} \tag{1.10}$$

$$U_{ij0}(0,1) = \underbrace{X_{i0}\omega_{0j}^{(0,1)} + \alpha_{ij0}^{(0,1)}}_{V_{ij0}^{(0,1)}} + \varepsilon_{ij0}^{(0,1)} \tag{1.11}$$

$$U_{ij0}(1,0) = \underbrace{X_{i0}\omega_{0j}^{(1,0)} + \alpha_{ij0}^{(1,0)}}_{V_{ij0}^{(1,0)}} + \varepsilon_{ij0}^{(1,0)} \tag{1.12}$$

$$U_{ij0}(11) = \underbrace{X_{i0}\omega_{0j}^{(1,1)} + \alpha_{ij0}^{(1,1)}}_{V_{ij0}^{(1,1)}} + \varepsilon_{ij0}^{(1,1)} \tag{1.13}$$

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. 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.

[18] 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.

[19] 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:

$$(\alpha_{m,}^{drink}, \alpha_{m,}^{mar}, \alpha_{m,}^{coc}, \alpha_{m,0}^{drink(01)}, \alpha_{m,0}^{drink(10)}, \alpha_{m,0}^{drink(11)}, \alpha_{m,0}^{mar(01)}, \alpha_{m,0}^{mar(10)}, \alpha_{m,0}^{mar(11)}, \alpha_{m,0}^{coc(01)}, \alpha_{m,0}^{coc(10)}, \alpha_{m,0}^{coc(11)})$$

1.4. Empirical Results

1.4.1. 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.[20]

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[21] (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[22]. 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.

Model B estimates positive structural correlation between drug-specific transitory shocks ($cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{mar}\right) = 0.58,\ cov\left(\varepsilon_{it}^{alcohol}, \varepsilon_{it}^{coc}\right) = 0.38,$ and $cov\left(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc}\right) = 0.46$).[23] 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.[24]

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.

[20] I do not report the parameter estimates of the initial conditions equation (equation 2).

[21] I define the order from softest to hardest, with alcohol a softer drug than marijuana, and marijuana a softer drug than cocaine.

[22] 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.

[23] 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 merely as a correlation between the time-varying unobserved component across drugs.

[24] 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.

The marginal effects[25] 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 average marginal effects (averaged over people and over time) reported in Table 3 should be interpreted only as an approximation. A more appropriate way to quantify the role of state dependence and stepping-stone effects is described in the next subsection.

1.4.2. 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,

$$P[Y_{ijt} = 1|Y_{i,j,t-1} = 1] - P[Y_{ijt} = 1|Y_{i,j,t-1} = 0] \quad [26], \tag{1.14}$$

under the assumption that state dependence is non-existent. In the absence of alcohol state dependence, the measure of alcohol persistence drops by 56%, from 54.55% to 23.57%, using the parameters estimated by Model A. Similarly, the measure of marijuana persistence drops by 57%, from 57% to 24%, in the absence of marijuana state dependence. Finally, the measure of cocaine persistence drops by 72%, from 38.84% to 10.81%, 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,

$$P\left[Y_{it}^{coc} = 1\middle|Y_{i,t-1}^{mar} = 1\right] - P\left[Y_{it}^{coc} = 1\middle|Y_{i,t-1}^{mar} = 0\right] \tag{1.15}$$

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 by 26%, from 15.4% to 11.37%, using the parameters estimated by Model A.

---

[25] These marginal effects for the logit specification and probit specification are written as follows

$$\frac{\partial P_{ijt}}{\partial Z_j} = f\left(V_{ijt}(\alpha_i)\right) * \beta_j = P_{ijt} * \left(1 - P_{ijt}\right) * \beta_j$$

$$\frac{\partial P_{ijt}}{\partial Z_j} = f\left(V_{ijt}(\alpha_i)\right) * \beta_j = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2[V_{ijt}(\alpha_i)]^2}\right) * \beta_j$$

where ϕ is the standard normal density

[26] 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. The difference in probabilities provided by equation 14 is driven partially by individual preferences and partially by state dependence.

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

$$P\left[Y_{it}^{coc}=1\middle|Y_{i,t-1}^{alc}=1\right]-P\left[Y_{it}^{coc}=1|Y_{i,t-1}^{alc}=0\right] \quad (1.16)$$

by 40%, from 6.78% to 4.66%, 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,

$$P\left[Y_{it}^{mar}=1\middle|Y_{i,t-1}^{alc}=1\right]-P\left[Y_{it}^{mar}=1|Y_{i,t-1}^{alc}=0\right], \quad (1.17)$$

by 18%, from 21.87% to 17.89%, 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).

1.4.3. Evaluating the Model

In order to assess the model's ability to predict drug use behavior, I evaluate 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.[27]

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.[28]

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

$$PCGF_j=\sum_{k=1}^{K_j}\frac{(O_{kj}-E_{kj})^2}{E_{kj}} \quad (1.18)$$

[27] 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

[28] In this version of the paper, I do not include the comparison between actual and predicted data for the logit model (Model A) with 3, 4, 5, and 6 mass points, because the best fitting logit model is achieved with seven mass points. I can provide the model fit tables upon request.

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.[29]

Table 2 reports the corresponding $PCGF_j$ for alcohol, marijuana, and cocaine. [30] Model C fits the data better than Model A, as measured by the log-likelihood, AIC, BIC, and PCGF.

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[31] in each period using parameters estimates from Model A. Columns 5 and 6 correspond to Model B, and columns 7 and 8 correspond to Model C.

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}[32] 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.

1.4.4. 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

[29] 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)

[30] The PCGF is calculated with different cells 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.

[31] 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).

[32] $Y_{ijt}, \varepsilon_{ijt}$ 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}

models depend on unobservable latent variables, they are approximated by the difference between the actual and the expected value, normalized for heteroskedasticity.[33]

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. [34].

$$E\left[r_{i,t}^{j}(\alpha_i)\right] = 0 \tag{1.19}$$

$$E\left[r_{i,t}^{j}(\alpha_i)^2\right] = 1 \tag{1.20}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5} \tag{1.21}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t}^{j\prime}(\alpha_i)\right] = 0 \tag{1.22}$$

$$E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j\prime}(\alpha_i)\right] = 0 \text{ for k=1,2,3,4,5} \tag{1.23}$$

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. [35] A model with second-order state 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

---

[33] Refer to the appendix for a description of how these generalized residuals were constructed, and refer to Gourieroux et al (1987) for a thorough description of generalized residuals for non-linear models. The generalized residuals are the difference between the actual binary variable Y_{ijt} and its expected value, normalized to correct for heteroskedasticity. 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}) $+ 1 * P\left(Y_{ijt}^{low}\right) + 2 * P(Y_{ijt}^{high})$, and
$var\left(Y_{ijt}\right) = P\left(Y_{ijt}^{none}\right) * (0 - E\left[Y_{ijt}\right])^2 + P\left(Y_{ijt}^{low}\right) * (1 - E\left[Y_{ijt}\right])^2 + P\left(Y_{ijt}^{high}\right) * (2 - E\left[Y_{ijt}\right])^2$. I did not report in the paper the generalized residual diagnostics for models with heterogeneous state dependence but I can provide them upon request.

[34] 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.

[35] 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.

different from zero. This is clear evidence of misspecification for the logistic models, but not for the probit specification (Model B). [36]

Finally, I show the cross-time, cross-drug correlation sample-analogue generalized residual $E\left[r_{i,t}^{j}(\alpha_i), r_{i,t-k}^{j'}(\alpha_i)\right]$ 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 to misspecification and hence it is not necessary to relax that assumption by incorporating higher order stepping-stone effects.

1.5 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), and by intensity of drug use (Model H).

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

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

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{age} age_{it}) \tag{1.24}$$

Table 8 reports that state dependence increases with age for all three drugs, while stepping-stone effects decrease with age[37]. 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})

---

[36] The correlation of the generalized residuals $E\left[r_{it}^{drink}(\alpha_i), r_{it}^{mar}(\alpha_i)\right] = 0.1759$, and $E\left[r_{it}^{drink}(\alpha_i), r_{it}^{coc}(\alpha_i)\right] = 0.0744$, 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 are positive and statistically significant, which is in line with the correlation of generalized residuals reported in Table 7

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

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.

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; and others), 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

1.5.2. Model E: Do Within and Between State Dependence Vary by Gender?

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

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{male} male) \tag{1.25}$$

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 not 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.

1.5.3. 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

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{\alpha_j} \alpha_j) \tag{1.26}$$

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}.

1.5.4. 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.

$$\gamma_{kj} = (\gamma_{kj}^{cons} + \gamma_{kj}^{\alpha_k} \alpha_k) \tag{1.27}$$

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 a 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}.

1.5.5. Model H: Does Within and Between State Dependence 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.[38]

[38] 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

The latent utilities for each drug now differentiate lagged low levels of use from lagged high levels of use.[39] The latent utilities for each drug j are as follows

$$U_{ijt} = \underbrace{\delta_{1j}(t-t_0) + X_{it}\beta_j + \sum_{k=1}^{J} \gamma_{kj}^{Low} Y_{i,k,t-1}^{Low} + \sum_{k=1}^{J} \gamma_{kj}^{High} Y_{i,k,t-1}^{High} + \alpha_{ij}}_{V_{ijt(\alpha_{ij})}} + \varepsilon_{ijt} \tag{1.28}$$

$$U_{ij0} = X_{i0}\,\lambda_{\,0j} + \alpha_{ij0} + \varepsilon_{ijt}^{initial} \tag{1.29}$$

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

$$Y_{ijt} = \begin{cases} 0, & U_{ijt} \le c_1 \\ 1, & c_1 \le U_{ijt} \le c_2 \\ 2, & U_{ijt} \ge c_2 \end{cases} \tag{1.30}$$

This model estimates the thresholds c1, and c2, 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 follow

$$\mathrm{P}\left(\mathrm{Y_{ijt}} = 0 \middle| \alpha_{\mathrm{ij}}\right) = \frac{\exp[\mathrm{c_1} - \mathrm{V}(\alpha_{\mathrm{ij}})]}{1+\exp[\mathrm{c_1} - \mathrm{V}(\alpha_{\mathrm{ij}})]} \tag{1.31}$$

$$\mathrm{P}\left(\mathrm{Y_{ijt}} = 1 \middle| \alpha_{\mathrm{ij}}\right) = \frac{\exp[\mathrm{c_2} - \mathrm{V}(\alpha_{\mathrm{ij}})]}{1+\exp[\mathrm{c_2} - \mathrm{V}(\alpha_{\mathrm{ij}})]} - \frac{\exp[\mathrm{c_1} - \mathrm{V}(\alpha_{\mathrm{ij}})]}{1+\exp[\mathrm{c_1} - \mathrm{V}(\alpha_{\mathrm{ij}})]} \tag{1.32}$$

$$\mathrm{P}\left(\mathrm{Y_{ijt}} = 2 \middle| \alpha_{\mathrm{j}}\right) = 1 - \frac{\exp[\mathrm{c_2} - \mathrm{V}(\alpha_{\mathrm{j}})]}{1+\exp[\mathrm{c_2} - \mathrm{V}(\alpha_{\mathrm{j}})]} \tag{1.33}$$

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. [40]

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.

---

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.

[39] 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.

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

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, high levels 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.

1.5.6. 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 while 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. This is in line with a finding that stepping-stone effects decrease with age.

While removing alcohol and marijuana use 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.[41] 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%[42], 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 to 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 cocaine 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 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.

---

[41] Carpenter and Dobkin (2009), for instance, presents evidence that a higher cost of alcohol accession (provided by the minimum legal drinking age) reduces binge drinking and heavy use of alcohol.

[42] 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 conterfactual 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.

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.

A second policy that decreases frequency and intensity of alcohol consumption among youth is the current minimum legal drinking age set at 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 was 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[43] 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 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.

1.6. 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

[43] 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.

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.

1.7. Appendix

1.7.1. Appendix A: Generalized residuals specification diagnostics

In this section, I describe step-by-step how to build the generalized residuals 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

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

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.

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 the following way

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

where

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

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

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

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

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

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, $\mathrm{P_{ijt}(\alpha_m)}$ corresponds to equation 3 for periods after 1998, and equation 4 for year 1998.

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

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

The type-specific correlation of generalized residuals are 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

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

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.

1.7.2. 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 or not 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, remains 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.[44] This is evidence that respondents for whom I do not have the entire sequence of drug-related questions for all three drugs were mainly lost 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[45], 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 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[46]), answers to sensitive questions were collected using audio computer-assisted self-interview (ACASI). ACASI

[44] 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.

[45] 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).

[46] 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.

interviewing system enabled the respondent to either 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 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. [47]

1.7.3. 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.

The 2010 report *Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings*[48] compares measures of lifetime and recent drug use patterns for the

[47] 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.

[48] Substance Abuse and Mental Health Services Administration, *Results from the 2010 National*

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 8^{th}, 10^{th} and 12^{th} 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 (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[49].

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

---

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.

[49] 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.

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 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.

2. The Effects of Alcohol Consumption on Cocaine Use, Substance Initiation and Criminal Participation: Regression Discontinuity Evidence from the National Longitudinal Study of Youth, 1997

2.1 Introduction

Because of the popularity of alcohol among youth, it is crucial for policy makers to understand the effect of alcohol consumption on other risky behaviors. Establishing causality between alcohol and criminal participation, for instance, is problematic, because there may be unobserved heterogeneity driving heavy alcohol consumption and other risky behaviors.

The key to disentangling the causal effect of alcohol consumption on other risky behaviors is to identify a research design that involves an exogenous variation in alcohol consumption. Previous studies have relied on different policies that exogenously change the cost of accessing alcohol to establish causality between alcohol consumption and risky behaviors. The policies that provide exogenous changes in the cost of accessing alcohol can be grouped into four main categories: (1) state-level changes in the minimum legal drinking age, (2) zero-tolerance laws, (3) college campuses that restrict alcohol use on campus, and (4) the minimum legal drinking age of age 21.

First, state-level change in the minimum legal drinking age occurred in the early eighties after the Reagan administration decided to increase the federal minimum legal drinking age from 18 to 21 in 1983. The minimum legal drinking age changed from 18 to 21 in different states at different time periods. By 1988, all states had implemented the new minimum legal drinking age of 21. DiNardo and Lemieux (2001) exploited the exogenous timing of implementation and found that increasing the minimum legal drinking age from 18 to 21 resulted in a reduction in alcohol consumption and an increase in marijuana consumption among high school seniors, indicating that marijuana and alcohol are substitutes.

A second source of exogenous change in the cost of alcohol accessibility is generated by Zero Tolerance Laws (underage drunk driving laws). The goal of these laws was to deter alcohol consumption among youth by suspending the license of drivers under the age of 21 who have any amount of alcohol in their blood. The widespread variation in the timing of adoption among states provided a clear identification strategy. Carpenter used this identification strategy to conclude that ZT laws decrease the probability of suicide (Carpenter 2004a) and heavy alcohol use (Carpenter 2004b) and to present evidence that heavy alcohol use causes the commission of property and nuisance crimes (Carpenter 2005). [50]

A third source of exogenous change in the cost of alcohol accessibility is generated by certain university's campus policies that forbid alcohol consumption on campus. Chaloupka et al (2004) relied on this identification strategy and found that alcohol and marijuana are complements.

Finally, a fourth source of exogenous change in the cost of accessing alcohol occurs at age 21, when individuals can start purchasing alcohol legally. Exploiting that discontinuity in alcohol consumption, Carpenter and Dobkin (2010) found that the increase in alcohol

[50] While Carpenter (2005) reaches different conclusion than this study, these differences can be reconciled by the difference in the age groups in our analyses. Carpenter (2005) focuses on an underage group, which is more vulnerable to alcohol consumption and more likely to drink irresponsibly relative to the adult group (age 21) that is examined in this paper.

consumption at age 21 results in an increase in arrests, mostly due to robberies, assaults, alcohol related offenses and nuisance crimes; Carpenter (2009) found that the increase in alcohol consumption at age 21 results in an increase in mortality, particularly in motor vehicle accidents and alcohol related deaths; and Yoruk and Yoruk (2011) found that recent marijuana consumption increased at age 21, while recent cigarette consumption remained unchanged.[51]

In this paper, I study the effect of alcohol consumption on three outcomes: (1) substance initiation, (2) cocaine consumption[52], and (3) self-reported criminal participation.

While Carpenter and Dobkin (2010) evaluate whether the probability of being arrested for a particular crime changes discontinuously at age 21, my study evaluates whether the share of respondents who report having engaged in a particular crime increased discontinuously at age 21. These outcomes would be equivalent if respondents always tell the truth, and if everyone arrested is actually guilty. Because these assumptions will most likely not hold, measures of self-reported crime participation are different from measures of arrest.

Identifying the causal effect relies on the exogenous decrease in the cost of accessing alcohol that occurs at age 21. While it is ideal in a regression discontinuity design to study outcomes in the month prior to the interview , reported outcomes in the year prior to the interview are also valuable. For instance, if respondents who are just older than 21 are much more likely to report having consumed a particular drug in the previous year than individuals who are just younger than 21, some of that discontinuous change in drug consumption must be driven by the exogenous increase in alcohol consumption that occurs at age 21.

This study finds that the decrease in the cost of accessing alcohol that occurs at age 21 results in the following: First, most measures of self-reported criminal participation remained unchanged at age 21 (violent attacks, property destruction, drug dealing), but the share of those who reported stealing in the last year decreased discontinuously. Second, alcohol consumption at age 21 decreases the probability of cocaine initiation and the number of users in the past year. These results are robust to different specifications.

The remainder of the paper proceeds as follows. Section 2 describes the National Longitudinal Study of Youth 1997 dataset that I use for this study. Section 3 describes the empirical strategy (regression discontinuity research design), and specifications for different models. Section 4 describes the results and robustness checks. Finally, section 5 offers a policy discussion and concludes.

2.2 Data

The data used in this analysis is extracted from the National Longitudinal Survey of Youth 1997 (NLSY97). This survey pertains to 13 waves corresponding to calendar years 1997 to 2009. The sample consists of 8984 individuals who were between the ages of 12 to 16 as of December 1996, and between the ages of 24 and 30 in the last available wave (2009). I restrict my analysis to the periods where adults are between the ages 19 and 23.

[51] Recent consumption of a particular drug is defined as consumption in the last month prior to the interview. The NLSY97 only reports the last month's consumption of alcohol, cigarettes and marijuana. Criminal participation and cocaine consumption are reported for the full year prior to the interview.

[52] Self-reported cocaine use last year answers to the question "Excluding marijuana and alcohol, since the date of 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 the public version of the NLSY97 does not contain information on the exact date of birth, it includes the month and year of birth as well as the age in months at the time of the interview.[53] The NLSY97 asks respondents questions regarding the consumption and intensity of use of a variety of drugs, including hard drugs, as well as participation in risky behaviors and criminal activity.

First, the NLSY97 asks participants questions related to alcohol participation in the last year (whether the individual reported having consumed alcohol since the date of the last interview), frequency of use in the last month (number of days the respondent had at least one alcoholic drink in the last month, conditional on having consumed alcohol since the last interview), intensity of use in the last month (number of drinks the respondent usually had on the days he or she consumed alcohol in the past month), and binge drinking in the last month (number of days the respondent had 5 or more drinks on one occasion, within a specified number of hours of each other, in the last month).

Second, the NLSY97 asks participants to report cocaine use in the last year (whether the respondent used cocaine since the last interview). To measure frequency of cocaine use, the survey asks respondents the number of times the respondent had used cocaine since the last interview.[54]

While the NLSY97 reports information on marijuana and tobacco frequency of use in the last month (number of days the respondent had used marijuana in the last month, conditional on having used it since the date of the last interview),[55] those outcomes were already examined by Yoruk and Yoruk (2011). This paper uses only measures of marijuana and tobacco initiation.

To use this data, I constructed measures of substance initiation using yearly information (whether the respondent used a particular drug[56] in the last year) accompanied by whether the respondent had used any of these drugs before the first wave. If the respondent had used drugs before the first wave, he or she is asked to report age of initiation. If not, initiation age can be inferred by looking at the wave during which drug use was first reported.

Finally, the NLSY97 contains measures of self-reported criminal activity in the last year, such as whether the individual has stolen anything, destroyed property, or attacked anyone with the purpose of hurting the victim, and whether the individual was involved in drug dealing. For a better understanding of how alcohol influences criminal participation, I also use measures of frequency of criminal participation conditional on participation.

During the first months after turning 21, measures of drug consumption or criminal participation in the last year include such behavior during the person's twentieth year. For that reason, there are advantages of measuring of alcohol consumption in the last month. However,

[53] Having age in months at the time of the interview allows me to construct the variable age_decimal at the time of the interview (age_decimal=age_month/12). For instance, individual with PUBID 94 is 232 months old at the time of the 1999 interview. Therefore, individual 94 will be assigned age_decimal 19.33 . The variable age_month is equivalent to the difference in months between date of birth and date of interview. Month and year of birth, as well as date, month and year of interview, are available in the public dataset.

[54] I will interpret the number of times the respondent consumed cocaine as the number of days of consumption in order to create a variable for the share of days last year when the individual consumed cocaine.

[55] Yoruk and Yoruk (2011) study the effect of the discontinuity in alcohol consumption which occurs at age 21 on marijuana and cigarette use. I use the marijuana variables to study the effect of alcohol consumption on marijuana initiation.

[56] I analyze initiation age for alcohol, marijuana, cigarettes, and hard drugs.

measures of cocaine use and criminal activity were reported at the yearly level. If respondents who are just older than 21 are drastically more likely to report consuming cocaine in the previous year than individuals who are just younger than 21, some of that discontinuous change in cocaine consumption must be driven by the exogenous increase in alcohol consumption that occurs at age 21.

2.3 Methods

To estimate the causal effect of alcohol consumption on hard drug consumption, I use a regression discontinuity research design that exploits the discontinuous decrease in the cost of accessing alcohol that occurs at age 21 due to the minimum legal drinking age.

I estimate the following reduced-form panel fixed effects model, to account for the longitudinal nature of the dataset. [57] I also cluster the standard errors at the individual level.

$$Y_{it} = X_{it}\gamma + \delta D_{it} + f^p(\widetilde{a_{it}}) + T_t + \alpha_i + \varepsilon_{it} \qquad (2.1)$$

In the previous equation, Y_{it} is the outcome of interest for individual i in year t, $D_{it} = 1[a \geq 21]$ is an indicator function of individual i being at least 21 years old at the date of interview in year t, and X_{it}represents a set of covariates for individual i at time t.[58] Some of the specifications I explore include an indicator for the month of the interview being the same as the individual ‘s birthday month, to control for the effect of birthday celebrations on drug consumption. The function $f^p(\widetilde{a_{it}})$ is a p-th order polynomial of an age-centered variable $\widetilde{a_{it}} = (a_{it} - 21)$ interacted with the indicator D_{it} , where

$$f^p(\widetilde{a_{it}}) = \beta_1\widetilde{a_{it}} + \cdots. + \beta_p\widetilde{a_{it}}^p + \beta_1^D\widetilde{a_{it}}D_{it} + \cdots \dots \dots \dots \beta_p^D\widetilde{a_{it}}^p D_{it} \qquad (2.2)$$

Due to the longitudinal nature of the NLSY97 dataset, I include a fixed effect α_i to account for time-invariant unobserved preferences across individuals. I also include year effects T_t.

The parameter δ identifies the causal effect of lowering the cost of accessing alcohol at age 21 on the outcome Y_{it} . The age polynomial is a function of normalized age (age centered at 21), which ensures that δ reflects the treatment effect on the outcome Y_{it} that occurs exactly at age 21. For the main part of the analysis, I estimate the model using first, second, and third order age polynomials on individuals between 19 and 23 years old.[59] Finally, ε_{it} represents the time-varying unobserved component.

The identification of the causal effect established by a regression discontinuity design is based on the assumption that other determinants of alcohol consumption, criminal participation, and hard drug use will transition smoothly before and after age 21. If that assumption holds, then the discontinuous jump in any particular risky behavior will be attributed to the discontinuous increase in alcohol consumption. Because individuals who are currently attending college may

---

[57] To make the coefficients readily interpretable, I report the coefficients of an OLS panel fixed effects model (xtreg) for all outcomes, whether binary or not (as opposed to logit estimates).

[58] A fixed effects model drops all time-invariant demographics, and only estimates coefficients for time-varying covariates. Because age at the time of the interview is already included in the age polynomials, and age was the only time-varying covariate, I do not include covariates in Model 1, 2 and 3. Only specifications in section E include an indicator for whether the interview takes place during the same month as the birthday, to control for the effects of birthday celebrations on alcohol consumption .

[59] While my dataset interviews individuals at age 12-16 in December 1996 and ends when individuals are between 24 and 31 years old in the last wave, I focus on individuals when they are within two years of age 21.

be most vulnerable to a change in the accessibility of alcohol[60], it is important to examine whether the probability of attending a four-year college transitions smoothly through the minimum legal drinking age of 21. I also examine whether there is non-random sorting of certain demographics at the threshold. Figure 1 shows graphical evidence of a smooth transition of demographics and college attendance through the threshold of age 21, which are consistent with the statistically insignificant δ reported in Table 2.

Another potential threat to this analysis is that individuals may become more likely to report alcohol consumption as soon as they turn 21, even if their alcohol consumption remained unchanged, because the outcome is illegal for individuals under 21. Carpenter and Dobkin (2009) show that there is a large increase in alcohol-related deaths and no change in reported lifetime drinking participation at age 21 (as opposed to recent alcohol consumption), which is compelling evidence that the increase in alcohol consumption that occurs at age 21 is unlikely driven by underreporting of alcohol use by individuals under the age of 21.

Given that demographics and college attendance transition smoothly through the threshold at age 21, I attribute the changes in risky behaviors that occur at age 21 to the exogenous decrease in the cost of alcohol consumption that occurs at age 21.

2.4. Results and Discussion

This section examines whether participation in risky behaviors changes in response to the sharp decrease in the cost of accessing alcohol that occurs at age 21. In particular, I examine whether self-reported criminal activity, consumption of hard drugs, and the probability of using other drugs for the first time change discontinuously at age 21. The first subsection documents the drastic increase in alcohol consumption that occurs at age 21.

2.4.1. Alcohol

The first stage of the regression discontinuity design evaluates whether there is a discontinuous increase in alcohol consumption in response to the decrease in cost of accessing alcohol that occurs exactly at age 21. I examine five outcomes: (1) alcohol use last year, (2) alcohol use last month, (3) binge use of alcohol in the last month (consumed 5 or more drinks in one occasion at least one day in the past month), (4) share of days on which alcohol was consumed in the last month (number of days in which the respondent used alcohol divided by 30), and (5) share of days on which 5 or more drinks were consumed in one occasion in the last month (number of days on which the respondent binged on alcohol divided by 30).

Each point in panels (A)-(C) of Figure 2 represents the proportion of respondents in each age bin that reports having consumed alcohol in the last year, consumed alcohol in the last month, or engaged in binge drinking in the last month (five or more drinks in one sitting in the same day), correspondingly. Panel (D)-(E) of Figure 2 represents measures of frequency (share of days drinking in the last month) and intensity of alcohol use (share of days binge drinking in the last month), correspondingly. I have eight bins on each side of the threshold of age 21, and they are determined as being eight-tiles, based on the age group. To quantify the size of the discontinuous increase in alcohol consumption that occurs at age 21, I present panel fixed effects regression

[60] Carpenter and Dobkin (2009) find that the overall increase in the mortality rate occurring at the discontinuity of age 21 is due to large increases in mortality among white males, particularly those currently attending college. Also, Williams et al (2004) report that the share of users of alcohol, marijuana and any illicit drug is higher among college students than among young adults in general.

estimates for each outcome. Table 3 presents the coefficient δ (equation 1) for models with different order centered-age polynomials interacted with an indicator that the respondent is older than 21.

Panel A of Figure 2 shows a discontinuous increase in the share of individuals who reported having consumed alcohol in the past year, from about 68% before turning 21 to near 74% at age 21. This discontinuous increase of 6 percentage points is statistically significant and robust to the order of the age-centered polynomial (Table 3).

Similarly, the proportion of people who reported consuming alcohol in the previous month increased by a large magnitude from 57% to 65% at age 21 (Panel B of Figure 2). This 8 percentage point increase is reflected in a statistically significant coefficient of 0.078 to 0.083 depending on the order of the polynomial.

The share of respondents who reported having binged on alcohol at least one day in the last month also increased discontinuously by 6 percentage points (Table 3), and this sharp increase is presented graphically in Panel C of Figure 2.

After establishing that all measures of alcohol participation increased sharply at age 21, I also examine measures of frequency of alcohol use. To measure frequency of alcohol use, Panel D and E of Figure 2 show the percentage of days on which alcohol was consumed in the last month (number of days of alcohol use in the last month, divided by 30) and the percentage of days on which the respondent reported binge drinking in the last month (number of days on which the respondent drank 5 or more drinks in one sitting, divided by 30), respectively. This increase in frequency of alcohol use is quantified in Table 3, where the coefficients indicate an increase of 3 percentage points in the share of days on which alcohol was consumed in the last month, and a 1.5 percentage point increase in the share of days on which the respondent reported binge drinking in the last month.

In summary, all measures of alcohol consumption, frequency, and intensity of use increased by a large magnitude at age 21.Given that the previous section presented evidence that demographics and college attendance transition smoothly through the age 21 threshold while alcohol consumption increases drastically at age 21, any change in the risky behaviors mentioned below will be attributed to the increase in alcohol consumption.

2.4.2. Cocaine

The NLSY97 asks respondents to report only two measures of cocaine consumption: (1) whether the respondent consumed cocaine since the date of the last interview (last year), and (2) the number of times the respondent consumed cocaine in the last year[61].

I analyze a measure of cocaine use (proportion of respondents who consumed cocaine in the last year) and a measure of frequency of cocaine use (percentage of days in the last year on which the respondent used cocaine). To allow for the possibility that alcohol has a heterogeneous effect on propensity to use cocaine versus its effect on the frequency of consumption among users, I analyze frequency of cocaine use separately for the entire sample, and for those who used cocaine at all last year.

[61] I interpret the number of times the respondent consumed cocaine since the last interview as the number of days of use since the last interview, which allows me to compute the percentage of days in the last year on which the individual consumed cocaine

Figure 3 (Panel A) illustrates a discontinuous decrease from 7% to 5.5% in the share of respondents who reported having consumed cocaine in the year prior to the interview that occurs at age 21. This decrease of 1.5 percentage points is statistically significant and consistent across specifications, order of age-centered polynomial, age bandwidth, and the inclusion of birthday effects.

While the number of cocaine users decreased, the frequency of cocaine consumption remains unchanged. Panel C of Figure 3 illustrates a slight increase in the frequency of cocaine use conditional on being a cocaine user (number of days on which cocaine was consumed last year, divided by 365, conditional on consuming cocaine in that year), but this increase is statistically significant only in certain specifications.

In summary, the number of respondents who report having consumed cocaine in the last year decreases at age 21, while the frequency of cocaine consumption remains unchanged.

2.4.3. Criminal Behavior

Previous literature found that the probability of being arrested increases drastically at age 21 (Carpenter and Dobkin, 2010). Measuring criminal participation using either arrest data or self-reported data would lead to the same results only if all respondents reported truthfully whether they had engaged in crime, if all arrestees actually committed the crimes for which they were arrested, and if everyone who committed a crime was arrested. As those are unlikely assumptions, measuring criminal participation with self-reported data will provide a different insight from measuring it with arrest data. [62]

This paper analyzes whether four measures of self-reported criminal participation change discontinuously at age 21: (1) stealing, (2) drug dealing, (3) attacking or hurting someone, and (4) destroying property. The NLSY97 asks respondents whether they have participated in any of those criminal activities in the past year. This study reports no evidence that criminal participation increases in response to the decrease in the cost of alcohol accessibility at age 21. In fact, there is a discontinuous and statistically significant decrease in the percentage of respondents who report having stolen since the last interview. These findings also hold when I restrict the sample to men only.

Stealing: The NLSY97 asks respondents separately whether they stole anything of value greater than $50 or of value under $50 since the last interview. I compute a measure of whether the individual stole anything since the last interview, regardless of the dollar amount. Panel A of figure 4 exhibits graphically that the proportion of respondents who report having stolen anything in the last year decreases discontinuously at age 21. This decrease of 2 percentage points is robust to the order of the age-centered polynomial, and to the inclusion of birthday effects. It also remains statistically significant when I restrict it to a balanced panel and when I explore an alternative age bandwidth.

[62] Brener et. Al (2003) discusses problems that arise with self-reported measures of risky behaviors and concludes that the computer-assisted personal interviewing (CAPI) system leads to more accurate reporting than in-person interviewing or self-administered questionnaires (pen-and-pencil- SAQ). The NLSY97 uses the CAPI system throughout the interviews. Sensitive questions, such as self-reported participation in risky behaviors, are administered via audio computer-assisted self-interview (ACASI) system which allows respondents to use headphones to listen to the questions, instead of reading them from the computer screen. http://www.bls.gov/nls/handbook/2005/nlshc2.pdf

<u>Selling Drugs:</u> The NLSY97 asks respondents whether they have sold illegal drugs since the last interview, and also asks separately whether the respondent has sold marijuana or hard illegal drugs. I focus on having sold illegal drugs in general, regardless of whether it is marijuana or hard drugs. Panel B of Figure 4 illustrates that the share of respondents who reported drug dealing since the last interview remains unchanged at age 21, which is consistent with the statistically insignificant estimates of such change reported in Panel B of Table 4.

<u>Attack or Hurt Somebody:</u> The NLSY97 asks respondents whether they have attacked anyone to hurt or fight since the last interview. Panel C of Figure 4 reflects that the share of respondents who were involved in a fight since the last interview does not change at age 21, which is consistent with the statistically insignificant estimate of such change reported by Panel B of Table 4.

<u>Destroy Property</u>: Finally, the last measure of criminal participation provided by the NLSY97 is an indicator for property destruction since the last interview. Once again, the probability of engaging in property destruction remains unchanged at age 21.

This study reports no evidence that self-reported criminal participation increases in response to the decrease in the cost of alcohol accessibility that occurs at age 21. In fact, there is a discontinuous and statistically significant decrease in the percentage of respondents who report having stolen since the last interview. Because the minimum legal age to be sentenced as an adult is 18, respondents have had enough time to adjust their criminal behavior to the law, and it is highly unlikely that reaching the legal drinking age would be driving any changes in criminal participation at age 21. This is in line with the percentage of arrests remaining unchanged as the respondents transitioned through the age threshold of 21.

2.4.4. Drug Use Initiation Age

To assess whether alcohol is a gateway drug, in the sense that alcohol consumption increases the probability of consuming other drugs for the first time, I analyze whether the probability of first time use of marijuana, cocaine, tobacco and alcohol itself change discontinuously at age 21. The first wave of the NLSY97 asks participants whether they have ever used alcohol, tobacco, marijuana and cocaine, and the age of initiation for each drug, conditional on having already used that drug by the first wave. For those who had not used a drug by the first wave and hence did not report initiation age in 1997, I impute starting age to be equivalent to the age at the time of the interview in which they first reported having consumed that drug.

Assigning drug initiation age is only possible for respondents with non-missing indicators of drug use for every period preceding the period during which the drug was used for the first time. For instance, respondent 2 and respondent 32 have the following sequence for alcohol use from years 1997 to 2009

R2=(y97=0, y98=0, y99=1, y00=0, y01=1, y02=1, y03=0, y04=1, y05=0, y06=. , y07=. , y08=1, y09=1)

R32=(y97=0, y98=0, y99=0, y00=. , y01=1, y02=1, y03=1, y04=1, y05=. , y06=. , y07=. , y08=., y09=.)

The imputed alcohol initiation age for respondent 2 corresponds to the age in year 1999, whereas the imputed alcohol initiation age for respondent 32 is unknown, because we do not know whether the first period of use was 2000 or 2001.[63] The average initiation age for tobacco is 15.04, for alcohol is 15.42, for marijuana is 16.61, and for cocaine is 18.27 (Panel B of Table 1). While the average starting age for all four drugs (alcohol, marijuana, cigarettes, and cocaine) are far below age 21, I examine whether the decrease in the cost of alcohol accessibility at age 21 increases the probability of trying other drugs for the first time among those who have not consumed it yet.

Alcohol: Panel A of Figure 5 presents the age profile of alcohol initiation. The probability of drinking alcohol for the first time decreases over time. Most individuals who will ever drink alcohol have already done so by age 21. Despite the continuous decrease in the probability of alcohol initiation, Figure 5 illustrates a discontinuous increase in the probability of alcohol initiation that occurs exactly at age 21. This 1 percentage point increase is statistically significant at the 5 or 10 percent significance level.

Tobacco: Panel B of Figure 5 presents graphical evidence of no change in the probability of tobacco initiation occurring at age 21, which is supported by the statistically insignificant estimates of such change.

Marijuana: Similarly, Panel C of Figure 5 presents graphical evidence of no change in the probability of marijuana initiation occurring at age 21, which is supported by the statistically insignificant estimates of such a change.

Cocaine: Finally, Panel D of Figure 5 reveals that a discontinuous decrease in the probability of cocaine initiation occurs exactly at age 21. Table 4 reports that the probability of consuming cocaine for the first time, among those who have not consumed it by age 21[64], decreases by 1 percentage point at age 21. This decrease is statistically significant at the conventional levels for first and second order age-centered polynomial, and statistically significant at the 10% significance level for the third order polynomial.

To summarize, the decrease in the cost of alcohol accessibility that occurs at age 21 results in an increase in the probability of alcohol initiation (among those who had not used it before), and a decrease in the probability of cocaine initiation (among those who had not used in before), while the probability of tobacco and marijuana initiation remain unchanged.

2.4.5. Robustness Checks

To diagnose the robustness of the previously reported coefficients, I estimate δ using alternative specifications. For this purpose, I already explored three models, with first, second, and third order age-centered polynomials, interacted with an indicator for whether the respondent is at least 21. To take advantage of the longitudinal nature of the dataset, I estimated a panel fixed effects regression model with year effects, and clustered the SE at the individual level. I did

---

[63] Panel B of Table 1 reports that about 60% of the sample were not lost to attrition and did not skip the drug-related questions. Because most of the missing data points are caused by respondents not answering the questionnaire at all for that period, rather than by respondents avoiding the questions related to drug use, these missing data points may be random. For periods in which data is missing, I assign zeros as the indicator for whether drug initiation occurred in that period, provided that drug use was already reported in earlier period.

[64] The probability of cocaine initiation is the number of first-time cocaine users in year t divided by the number of respondents who had not consumed cocaine as of year t-1. The probability of consuming cocaine at age 21 is different from the probability of initiating cocaine consumption at age 21.

not include demographic variables because a fixed effects model would drop them. These specifications used the entire panel dataset and periods when the respondents were between ages 19 to 23.

As a robustness check, I estimate the model with second order age-centered polynomial with three variations: First, all specifications in Table 5 include an indicator for whether the interview took place in the same month as the birthday of the respondent, to account for birthday celebration effects. Second, the first column extends the age bandwidth to 18-24. Third, the third and fourth columns restrict the analysis to respondents for whom I have a balanced panel.[65]

Table 5 confirms the results reported in the previous section. When I restrict the analysis to a balanced panel, all measures of alcohol consumption increase at age 21 (alcohol consumption last year, alcohol consumption last month, and binge drinking last month, as well as the percentage of days on which alcohol was used in the last month and last year). The same occurs when I increase the age bandwidth.

Panel B of Table 5 shows a statistically significant decrease in the share of respondents who reported using cocaine in the last year among those who just turned 21, and these estimates are robust to the specification, inclusion of birthday effects, and age bandwidth. While the decrease in the number of cocaine users is clear, the frequency of cocaine consumption remains unchanged. There is limited evidence of an increase in the frequency of cocaine use last year when I restrict it to those who consumed cocaine last year. However, this increase is not statistically significant for all specifications.

The probability of drinking alcohol for the first time increases at age 21, and this estimate is robust to various specifications as well. The decrease in the probability of cocaine initiation is robust to inclusion of birthday effects, and to restriction to observations with a balanced panel. While the coefficient remains negative when I use an alternative age bandwidth (18-24), it is not statistically significant in the conventional sense. The probability of initiation to tobacco and marijuana remained unchanged at age 21.

Regarding self-reported criminal activity, the only crime that exhibits a statistically significant decrease at age 21 is stealing. There is no change in the share of respondents who reported selling illegal drugs, attacking someone with the purpose of hurting the victim or destroying property. These results are robust to inclusion of birthday effects, change in the age bandwidth, and using a balanced panel. Consequently, the share of respondents who reported having been arrested in the last year remains unchanged at age 21.[66]

2.5. Conclusions and Policy Implications

This paper examines the role of an increase in alcohol consumption at age 21, generated by the minimum legal drinking age, on risky behaviors. In particular, I analyze whether certain risky behaviors, such as cocaine consumption, participation in criminal activity (stealing, drug dealing, attacking somebody, and property destruction), and drug initiation change discontinuously at age 21 in response to the ease of alcohol accessibility.

[65] The balanced panel is restricted to those whom I can observe for 4 or 6 periods when the age bandwidth is 19-23 or 18-24 respectively, with at least one observed period on each side of the age threshold 21.
[66] A previous version of the paper reported that the share of respondents who reported having been arrested in the last year remained unchanged at age 21.

The minimum drinking age law of 21 affects all measures of alcohol consumption for individuals just under 21 years of age. The share of individuals consuming alcohol last year, and last month, as well as the percentage of respondents who reported binge drinking last month, increased at age 21. Alcohol intensity of use, measured by the percentage of days of drinking and binge drinking last month, increased by 4 and 1 percentage points respectively. The increase in alcohol consumption and intensity of use that occur at age 21 are statistically significant. While most individuals have already used alcohol before age 21, the probability of alcohol initiation is also 1.5 percentage point higher for those just above age 21 relative to those just below age 21.

The results of this paper can be summarized as follows:

First, the number of cocaine users in the last year decreased discontinuously at age 21 by 1.8 percentage points, and this effect remains robust across different specifications.

Second, there is no evidence of an increase in self-reported criminal activity at age 21. In fact, the percentage of respondents who reported having stolen anything in the last year decreased at age 21 by 2 percentage points, and that decrease was statistically significant across specifications. These results also hold when I restrict the analysis to men only.[67] The other self-reported measures of criminal participation remained unchanged at age 21.

Third, this study presents evidence that, at age 21, alcohol is not a gateway drug, defined as alcohol increasing the probability of initiation into other drugs. While the probability of cigarette and marijuana initiation transitions smoothly over the age of 21, the probability of consuming cocaine for the first time decreases at age 21.

I want to emphasize that these findings are specific to the effects of alcohol exactly at age 21, without any intention of generalizing these findings to other age groups. Most respondents who initiate use of hard drugs do so at an earlier age than 21. Hence, respondents who initiate cocaine use between age 21 and 22 are not a representative sample of the population of hard drug users.

While lowering the cost of accessing alcohol may have negative consequences in youth mortality (Carpenter and Dobkin 2009), or increase the probability of recent marijuana consumption among users (Yoruk and Yoruk 2011), this study finds that it lowers the probability of cocaine use, the probability of initiation into hard drugs, and the probability of stealing, while leaving unchanged the probability of participating in property destruction and attacking someone with the intent to hurt the victim.

[67] The results for criminal participation remained unchanged after I restrict the analysis to men only. This can be attributed to the fact that most women never participate in criminal activities, and so their sequence of criminal participation was unchanged. Thus, they were dropped when I used the fixed effects model.

3. Is Three Strikes Law Crowding the Courts? Evidence from the State Court Processing Statistics 1990-2006

3.1 Introduction

Between 1993 and 1995, several states enacted *Three Strikes, You're Out* Sentencing Laws (TSL) with the goal of removing violent repeat offenders from society for longer periods of time, even though most states already had enhanced sentencing for repeat offenders.[68] Because the ultimate goal of implementing Three Strikes Laws (TSL) was to reduce crime, economists and policy makers have been interested in measuring the effectiveness of the passage of Three Strikes Laws in decreasing recidivism, either by incapacitation or deterrence (Iyengar, 2012; Shepherd 2002), ignoring how this law affected other aspects of the justice system.

Studying the effects of sentence enhancement provided by TSL on other aspects of the criminal justice system is potentially important for policy. If TSL provides an incentive for potential strikers (individuals arrested for a crime defined as a potential strike[69]) to fight the conviction charge in trial, as opposed to pleading guilty, then case processes would be delayed, and court rooms and jails would be overcrowded. While such an increase in cases resolved by trial is clearly costly to the state, opponents of plea bargains may find it to be a positive unintended consequence that TSL encourages defendants to exercise their rights to a criminal trial.

Previous studies provide evidence that the threat of "tough sentences" (i.e. the death penalty) affects the action of defendants when faced with the decision to accept plea bargains versus going to trial (Kuziemko 2006). In particular, Kuziemko (2006) finds that the passage of death penalty laws in New York encouraged defendants to plea bargain with harsher terms (to their original charge as opposed to a lesser sentence), without affecting the propensity to pleading guilty altogether, in order to avoid death row.

While defendants who are faced with the threat of life imprisonment (potential third strikers) should have similar incentives to pleading guilty as those facing death row, it is unclear whether potential first or second strikers would alter their behavior towards pleading guilty. While in most TSL states (states that have adopted Three Strikes Laws) the sentence length remains unaffected at the first two strikes, the label of being a striker affects the future criminal career of the defendant. TSL provides an incentive to forward-looking defendants to fight in trial the possibility of being convicted for a strike.

When studying the effects of policies such as TSL that attempt to influence criminal behavior, a common concern is that such studies require the assumption that offenders are informed about the current laws, and that they make rational decisions taking into account costs and benefits of their criminal actions. While many potential criminals may not be aware of the sentencing laws in their state at the time they commit a crime, they are more likely to be informed when they are deciding whether to demand a trial or plead guilty, because they are provided information by their defense attorney. Furthermore, while most crimes are committed

---

68 Bureau of Justice Assistance, *National Assessment of Structured Sentencing*, U.S. Department of Justice, February 1996.

69 I define a potential strike as a violent crime, since those are commonly covered in the category of potential strikes, regardless of the state and number of prior strikes.

on the spur of the moment, or while intoxicated, criminals are at their most rational and less likely to be under the influence of drugs or alcohol after arrest, when being advised by their defense attorney, and may even be held in jail while deciding whether to demand a trial.

To the best of my knowledge, Bjerk (2005) is the only study that analyzes the effects of Three Strikes on pre-trial outcomes; that study finds that prosecutors are more likely to reduce a defendant's prosecution charge to a misdemeanor when the arrest charge makes them potential third strikers.

Relative to the existing literature on the unintended consequences on three strikes laws, I contribute by providing empirical evidence that defendants are forward-looking, and react to the three strikes law by going to trial instead of pleading guilty[70], in order to avoid being convicted with crime that is in the strike category, even in cases when TSL does not enhance punishment for the current strike. Because of the cost of having these additional cases resolved in trial, my study shows the importance of accounting for this type of pre-trial decision when estimating the total cost of Three Strikes.

I use data from the 1990-2006 State Court Processing Statistics database and a state-by-year difference-in-differences research design to compare the change in the likelihood of plea bargaining by violent offenders after the passage of Three Strikes laws, relative to the trend among non-violent offenders. While the definition of what constitutes a strike varies slightly by state and by number of previous strikes, violent crimes are always considered a strike (Clark et. al 1997). [71]

A limitation of the dataset is that the crimes included in the "strike zone" are more specific than the categories of crime provided in the dataset. For instance, California includes drug sales to minors in the "strike zone" (Clark et. al 1997), but I cannot distinguish in the data whether the defendant was arrested for drug sales to adults or minors. This measurement error affects who is classified as a potential striker versus a defendant who is arrested for a felony but is not facing the threat of being convicted for a strike. This measurement error will understate the true difference between the treatment and the control groups.

My empirical work employs a difference-in-difference estimation strategy, separately for defendants with at least one prior violent conviction and defendants with no prior violent convictions.[72] Separating the analysis by whether the defendant has prior violent convictions allows me to analyze how the defendant responds to TSL when threatened by the severity of the

[70] It is conventional in the law and economics literature on plea bargaining to assume that a plea bargain is less costly to the prosecutor than trial, and that filed cases are settled either with trial or plea bargains (Oren and Oren 2004). My dataset indicates whether a case was resolved with plea bargaining, but it does not indicate explicitly whether it proceeded to trial. Following the literature, I assume that cases that are not secured through plea bargaining proceed to trial.

[71] The most serious arrest charges can be categorized in four groups: violent, property, drug-related, and public order crimes. Violent crimes are subdivided into murder, rape, robbery, assault, and other violent offenses. Among non-violent felonies, the property crimes are subdivided into burglary, larceny-theft, motor vehicle theft, forgery, fraud, other property crime. Drug-related crimes are subdivided into drug sales, and other drug-related crimes. Finally, public order crimes are divided intp weapons, driving-related, and other public order crimes.

[72] While the dataset reports the actual number of prior convictions, it only reports whether the respondent has had at least one prior violent conviction, and not the actual number of prior violent convictions (variable is named "privconv" in the dataset and was recoded to be an indicator). Therefore, I cannot clearly separate potential second from third strikers.

current punishment (for later strikes) and when only threatened with regard to their future criminal path and not by the severity of the current sentence (for earlier strikes).

Furthermore, I perform this analysis on three subsamples: California only, states that implemented TSL excluding California, and finally states that never implemented TSL. I analyze California separately for four reasons: First, California requires only the first two convictions to be offenses in the "strike zone," and then any subsequent felony (violent or not) counts as a third strike. Second, California has a Two-Strike law in addition to the Three Strikes Law, which doubles sentence length for any felony, whether violent or not, for those who have previously been convicted for a strike. Third, in California, a third striker has the possibility of parole after serving a minimum imprisonment of 25 years, as opposed to other states that sentence the defendant on life imprisonment without parole (Clark et al 1997). Fourth, California legislation officially allows prosecutors to drop previous felony convictions or strikes "in the furtherance of justice[73]" (Bjerk, 2005). Because California comprises 23% of the sample, I analyze TSL states excluding California to assess whether the estimates attributed to TSL are simply picking up effects from California. Finally, performing the analysis on states that never implemented TSL provides evidence that the effects attributed to TSL were not driven by a third factor that changed the behavior of violent offenders around the time that TSL was implemented.

My study finds that, in California, defendants facing the threat of being convicted for a strike are 3 to 4 percentage points less likely to plead guilty after the passage of Three Strikes Laws. This effect is statistically significant among those who are facing their potential first strike, even though the length of incarceration for the first violent conviction remains unaffected after the implementation of TSL. This is evidence that defendants are forward-looking and take their future criminal career path into account when making decisions. Because the two-strike law in California increases sentence length for defendants who are arrested for a felony (violent or non-violent) if they have a previous violent conviction, and the control group is composed of defendants arrested for non-violent felonies, I did not expect those arrested for a violent felony to react differentially to those arrested for a non-violent felony.

States other than California that also have a two-strike law[74] require both convictions to be from the "strike zone." Among those facing the threat of being convicted for their second strike, I expect those arrested for a violent felony to have a greater incentive to fight the charge at trial than those arrested for a non-violent felony. Among those facing the threat of being convicted for a third strike, all states, except for Utah (which only comprises 1.08% of the sample) require two prior convictions from the "strike zone." While I cannot separate potential second from potential third strikers in my sample[75], I expect them to react differentially from defendants with a prior strike who are arrested for a non-violent felony. Excluding California, defendants facing the threat of being convicted for a second/third strike are 8 to 9 percentage points less likely to plead guilty following TSL than defendants who are not facing the same threat.

[73] California Penal Code 667(f)(2). When the prosecutor drops a previous charge or strike, the charge is not dropped altogether, but it does not count against the defendant for the current charge (Bjerk 2005). See Walsh (1999) for a discussion of criteria used to apply such discretion with regard to previous charges.
[74] See Table 1. For instance, Georgia, Pennsylvania, and Tennessee also increase the punishment for those convicted for a second strike, but they require that both the first and second strike are offenses from the "strike zone."
[75] Potential second/third strikers are defendants with at least one prior violent conviction who are currently arrested for a violent offense.

These effects are robust to the inclusion of several characteristics, and to a difference-in-difference-in-difference (DDD) estimation, indicating that a potential second/third striker in TSL states is 4-6 percentage points less likely to plead guilty than a "potential second/third striker" in a state where TSL laws were never sought, and this effect is statistically significant among those with at least one prior violent conviction.

This paper is organized as follows. The next section discusses the data, while section 3 presents a basic model to illustrate the decision process of the defendant faced with the threat of being convicted for a strike. Section 4 describes the empirical approach, and the different specification of a difference-in-difference approach. Finally, section 5 summarizes and concludes.

3.2 Data

I use data from the "State Court Processing Statistics 1990-2006: Felony Defendants in Large Urban Counties."[76] This individual level dataset is collected by the United States Department of Justice with the ultimate goal of determining whether local level data was collected and aggregated accurately at the national level. This cross-sectional dataset collects information about defendants who were arrested for state felony offenses in the month of May, biannually from 1990 to 2006, and tracks them until final disposition, or for a full year after the date the case was filed.

The dataset is composed of 134,767 cases. It includes the year and state where the arrest took place, the final adjudication of the case (whether the defendant was convicted), the type of adjudication (trial, dismissal of case, guilty plea), demographics (age, race, gender), criminal history (number of prior arrests, incarcerations, convictions, misdemeanor convictions, felony convictions, and an indicator for whether the defendant has at least one prior violent conviction), and details about current arrest (most serious arrest charge, number or arrest charges), among other variables. Therefore, the State Court Processing Statistics dataset is nearly ideal for studying whether defendants who are now at risk of facing more lengthy sentences (potential strikers, as defined by prior violent convictions) would now have an incentive to demand trials, hence crowding courts and jails, and imposing an unintended cost in the justice system.

From the 134, 767 cases, I restrict my analysis to not pending or missing cases (with a valid adjudication outcome, or conviction status[77]). Among the remaining 118,228 cases, I further restrict my analysis to those with valid data for the most serious conviction charge and most serious conviction category[78]. At this point, I have 118,061 observations remaining, and I delete those with missing level of most serious adjudication charge[79], regardless of whether this data is missing because the case is pending or because the variable itself is missing. Finally, I will

[76] The dataset is collected by the United Stated Department of Justice. It can be downloaded from the Inter-university Consortium for Political and Social Research (ICPSR), and corresponds to study 2038.

[77] I delete observations where the adjudication outcome (variable is named adj1) is pending or missing, which leaves me with potential adjudication outcomes of convicted, not convicted, or other outcome.

[78]The most serious conviction charge and category (variables are convoff and convtype, respectively) indicate whether the most serious conviction charge was violent, property, drug, public-order , other felony, misdemeanor, or not applicable. Not applicable corresponds to those who were not convicted. I erase the missing values, not the "not applicable" cases.

[79] The most serious level of adjudication charge (variable is named achglevl) that are non-missing can be categorized as either felony or misdemeanor.

restrict the remaining 115,571 to those cases where the type of adjudication[80] is non-missing, and hence I can tell whether the case was dismissed, acquitted, diverted, resolved through plea bargaining, convicted through trial, or other.[81] These restrictions result in 114, 363 cases, of which only 97,215 have non-missing age, gender, and race. This subsample of 97,215 cases is the subsample that I use for this analysis

Table 2 presents a summary of the characteristics of this subsample by whether the arrest occurred in a state that implemented TSL between 1993 and 1995 and by whether the defendant is currently arrested for a violent or non-violent felony.

The differences across subsamples presented in Table 2 can be summarized as follows: First, blacks are more likely than whites to be arrested for a violent offense, and this pattern holds in states with and without TSL. Second, defendants arrested for a non-violent offense are more likely to have prior convictions, whether for felonies or misdemeanors, than those arrested for a violent offense, and this pattern holds in states with or without TSL. This can be reconciled if those who are currently arrested for a violent offense are more likely to have been convicted for violent offenses already (which we can see it is true in states with and without TSL), and hence may already be imprisoned. Finally, more than 90% of the convictions were resolved through plea bargaining, and this percentage is even higher among those arrested for a non-violent offense.

The State Court Processing Statistics provides weights, which can be used to make the sample representative of all “felony filings during the month of May in the 75 most populous counties” where these counties “account for more than a third of the United States population and approximately half of all reported crimes.” [82]

3.3. Decision Process Following Three Strikes

This section illustrates the decision process of the defendant, following TSL. I start the decision process with a defendant who is currently arrested for a violent offense, which can potentially result in a conviction for a first strike. [83]

3.3.1. TSL States Other than California

In most TSL states other than California, TSL mandates life imprisonment on the third conviction only if all three convictions are specified violent felonies,[84] without affecting sentence length for non-violent convictions, or for the first or second violent conviction.

---

[80] The type of adjudication (variable named adjtype) after deleting the missing observations leaves me with potential outcomes of dismissal, acquittal, divert-defer, guilty-plea, guilty-trial, and other.

[81] Convictions are reached through pleading guilty or trials. Defendants who were not convicted either had their case dismissed or were acquitted. If the case is resolved through a guilty plea, then the defendant is convicted with certainty. On the other hand, if the defendant does not plead guilty, the case does not necessarily end up in conviction. Among non-dismissed cases that were not resolved through pleading guilty, the outcomes were conviction in trial, acquittal, or other (over 90% of “other” ended up in divert-defer).

[82] U.S. Dept. of Justice, Bureau of Justice Statistics. STATE COURT PROCESSING STATISTICS, 1990-2006: FELONY DEFENDANTS IN LARGE URBAN COUNTIES [Computer file]. Conducted by Pretrial Justice Institute (formerly, the Pretrial Services Resource Center) [producer]. ICPSR02038-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010.

[83] While TSL could clearly have an impact on the probability of committing an offense which could result in a first strike, I ignore the decision node where the defendant does not have any prior strikes and is faced with the decision about committing crimes. Those individuals will never be observed in the dataset and will not be affected by TSL.

I assume that every defendant who is arrested for a violent offense and whose case was not dismissed will be offered the possibility of pleading guilty[85] to a violent offense only[86]. I assume that anybody who commits a crime is arrested for that crime[87].

Figure 1 graphically illustrates the decision process of a potential criminal who has no previous strikes, but is currently arrested for a violent offense (state (s)), and hence is facing the threat of being convicted, creating a first strike. This violent offense is punishable by T_1 years in prison (whether or not this state incorporated TSL)

At this point, the defendant has two choices: He can plead guilty to a violent offense, and the plea bargain will lower his sentence by γ_1 years. After serving $T_1 - \gamma_1$ in prison, he will be out of jail in state S (having one prior strike). Therefore, the payoff to pleading guilty at stage (s) is the disutility from being in prison for $T_1 - \gamma_1$ years plus the continuation value of reaching state S.

If defendant does not plead guilty, he will be convicted, creating a strike, incarcerated for T_1 years, and released in state S, with probability π_1. With probability (1-π_1), the defendants will be convicted of a crime outside the strike-zone (or acquitted, even though this is unlikely[88]), in which case he will be incarcerated for a number of years NS_1 ($NS_1 < T_1 - \gamma_1$), and move back to original state N to face the decision to commit a crime.

TSL did not affect the sentence length T_1 at the first strike, but only the continuation value of being in state S relative to N. Figure 2 illustrates the future path of those who moved to state S.

If this defendant is now in state S (Figure 2), he faces the decision to commit a misdemeanor (m), non-violent felony (f), or violent felony (s). Each choice includes in its payoff the instant utility of the crime and a continuation value. The effects of TSL on recidivism are beyond the scope of this study, and I ignore the decision to commit a crime that is not in the strike zone. A defendant facing the threat of being convicted for his second strike (state S(s)) maximizes the expected value of lifetime utility by choosing whether to plead guilty.

The committed crime is punishable by T_2 years of prison; that sentence length is unaffected by TSL. If the defendant pleads guilty, the sentence is reduced by $\gamma_2$years, but he will move to state SS after serving in prison. Therefore, the payoff from pleading guilty when in state S(s) is the utility from serving $T_2 - \gamma_2$ years and the continuation value in state SS. If he chooses not to plead guilty, then with probability π_2 he will go to prison for T_2 years and be released with state SS, and with probability $(1 - \pi_2)$ he will go to prison for a lesser conviction (incarcerated for NS_2 years) and then released with state S.

---

[84] WI, WA, VA, TN, NJ mandate life without parole for third conviction for certain violent felonies. GA and IN mandate life imprisonment for a second specified violent felony conviction.

[85] It is reasonable to assume that every defendant is offered a guilty plea, because over 90% of convictions are obtained by guilty pleas (Grossman and Katz 1983; Oren and Oren 2004; also see Table 2).

[86]It is reasonable to assume that those arrested for a violent offense are offered the opportunity to plead guilty only to a violent offense. Panel B of Table 2 shows that, in TSL states, 70% of convicted defendants who pleaded guilty and were arrested for a violent offense are convicted for a violent offense. That does not imply that they were offered a guilty plea to the original charge. For instance, somebody arrested for murder can be offered a chance to plead guilty to manslaughter, which will entitle him to fewer years of prison, but will still label him as a "first striker." Furthermore, Table 8 shows that the share of cases that resulted in a plea bargain to a violent conviction among those arrested for a potential strike remained unchanged relative to those arrested for a non-violent felony, following the implementation of TSL. California had a different effect. TSL decreased the share of plea bargains for a violent offense among potential strikers. See Table 8.

[87] This is a strong assumption, but, from the viewpoint of the researcher, those who are never arrested for their committed crime will never appear in the data.

[88] Table 1, column 1 indicates that potential strikers in California are convicted in 82% of the cases.

Once again, TSL did not affect the sentence length T_2 but only the continuation value of being in state SS. Figure 3 illustrates the future path of those who moved to state SS.

Again, I ignore the decision to commit a crime that is not in the strike zone. A defendant facing the threat of being convicted for his third strike (state SS(s)) maximizes the expected value of lifetime utility by choosing whether to plead guilty. In the absence of TSL, this violent crime would have been punishable by T_3years of prison. Following the implementation of TSL, the sentence upon conviction for a third violent offense corresponds to life imprisonment.

If he pleads guilty to a violent conviction, his sentence may be lowered by γ_3, and the payoff is the utility from serving somewhat less than life imprisonment and the continuation value from being released at a very old age. If the defendant chooses not to plead guilty, then with probability π_3 he will be imprisoned for life, reaching a final state SSS. On the other hand, with probability $(1-\pi_3)$ he will go to prison for a lesser conviction (incarcerated for NS_3 years) and then will be released with state SS.[89]

While defendants arrested for a violent conviction always had an incentive to fight the charge, hoping to be convicted of a lesser charge, TSL changes the incentives in the following ways: First, because TSL severely enhances the punishment for a third violent conviction, it increases the incentive to risk a trial as opposed to pleading guilty to a strike (unless γ_3 is very large). Second, because TSL increases the punishment only for those who have two prior violent convictions, it decreases the continuation value at any state that includes a conviction for a strike. Because the value of being in a state such as SS is lower following TSL, I expect an increase in the share of cases that now have an incentive to risk a trial. In this case, we would expect the treatment group (arrested for violent conviction) to react differentially from the control group (arrested for non-violent felony) at every potential strike.

3.3.2. California

The California legislation defines a punishment for a defendant convicted for a strike as "mandatory doubling of sentence for any felony if one prior serious or violent felony conviction; mandatory life without parole for 25 years for any third felony conviction if two prior serious or violent-felony convictions." (Clark et al 1997).

In the case of California, I expect the treatment group (arrested for violent conviction) to react differentially from the control group (arrested for non-violent felony) at the potential first strike (Decision tree is the same as in figure 1). While there is no prison enhancement associated with being convicted with a violent or non-violent felony, as long as the defendant has no prior violent convictions, the treatment group now faces the threat of being released with a label of a "first-striker," which has a lower continuation value following the implementation of TSL.

Among defendants with one prior violent conviction, both control and treatment group face the threat of a longer imprisonment, following the implementation of TSL. Figure 4 illustrates how the two-strike law affects defendants arrested for a potential second strike. Among defendants with two prior violent convictions, both control and treatment groups face the threat of being sentenced to life imprisonment. Following TSL, I expect the treatment and control

[89] I consider SSS a terminal state, whether the defendant reached that state by trial and is now in life imprisonment, or by plea bargain, where he obtained a conviction somewhat less than life imprisonment. The effects of TSL on potential third strikers are ambiguous. The threat of facing life imprisonment may deter the defendant from going to trial only when γ_3 is large enough. Unfortunately, I cannot distinguish in the data whether the defendant was arrested for a potential second or third strike.

group to adjust their behavior in accordance with the increase in the total payoff from risking a trial versus accepting a plea bargain.

3.4. Empirical Estimation

The goal of this study is to analyze whether Three Strikes Laws decrease the probability of resolving a case through plea bargaining for those who were arrested for a "potential strike" (violent offense) relative to those who were arrested for a non-violent offense.

I employ a difference-in-difference estimation strategy, where the treatment group consists of defendants arrested for a violent offense, and I run separate estimations for those with and without previous violent convictions. This gives an insight into the role of option value when choosing to fight a potential strike by going to trial because, at the first strike, (and in most states also at the second strike) the sentence length is the same whether or not the state has incorporated a Three Strikes Law.

The dependent variable $Plea_{it}$ is the probability that the non-dismissed case i filed in year t was resolved by pleading guilty. Table 3 shows the proportion of felony defendants older than 18 with non-dismissed cases whose case was resolved through plea bargaining. The proportion of defendants arrested for a violent offense who pleaded guilty fell from 85.7% to 83%. On the other hand, the proportion of defendants who were arrested for non-violent felonies who pleaded guilty fell only slightly, from 88% to 87%. Relative to those arrested for non-violent felonies, defendants arrested for violent felonies are 1.7[90] percentage points less likely to plead guilty, following the implementation of TSL. This effect is larger in California.

3.4.1. California

I analyze California separately from other states that passed Three Strikes Laws for reasons mentioned previously.

I estimate the following equation restricted to cases filed in California only, where $Plea_{it}$ is defined above. *After* is a dummy variables coded as one if the defendant's case was filed after 1994 (the year in which Three Strikes was enacted in California) and zero otherwise. *Treat* is a dummy variable coded as one if the defendant was arrested for a violent offense.

$$Plea_{it} = \alpha + \beta_1 After_t + \beta_2 Treat_i + \beta_3 After_t * Treat_i + \gamma X_i + \tau_t + \varepsilon_{ijt} \tag{3.1}$$

Table 4 can be summarized as follows. In California, following the implementation of the Three Strikes Law, the probability that a potential striker (those arrested for a violent offense) pleaded guilty decreased by 3 to 4.6 percentage points, relative to those arrested for non-violent offenses,. For potential first strikers, the probability of pleading guilty decreased by 2.6 to 4 percentage points. Specifications (4) and (5) are statistically significant in the conventional way, and specifications (2)-(3) are statistically significant at the 10% significance level for adults.

These estimates are statistically significant and robust to the inclusion of year effects (τ_t), demographics, criminal history, and characteristics of the current arrest (X_i)[91]. Also, these

[90] This is statistically significant at the 10% significance level.

[91] Demographic variables include gender, race, and age. Characteristics of the current arrest includes the total number of arrest charges, as well as an indicator for the most serious arrest charge, which includes 16 categories. Criminal history includes number of prior convictions, number of prior felony convictions, number of prior misdemeanor convictions, number of previous imprisonments, number of previous times in jail, and an indicator for having been convicted, convicted for felony, convicted for a misdemeanor, or incarcerated at least once before.

results remain robust when I restrict them to defendants who were at least 18 years old when the case was filed. Table 4 reports the marginal effects of a probit estimation, with robust standard errors.

In California, those who already have one prior violent conviction (one strike), can become a second or third striker with a felony regardless of whether it is violent or non-violent. Because the treatment group is composed of those who are arrested for a violent felony and the control group is composed of those arrested for a non-violent felony, both groups should respond similarly to the implementation of Three Strikes Laws.[92]

3.4.2. All States That Passed Three Strikes Laws

Equation (2) estimates a difference-in-difference estimation, similar to equation (1), but it includes cases filed in states that have passed Three Strikes Laws other than California. The coefficient of interest is α_3, which estimates whether the probability of pleading guilty changes in response to the implementation of Three Strikes Laws among potential strikers relative to those arrested for non-violent offenses,

Because California comprises 23% of the sample and has a two-strike law in addition to the three-strike law, I separately estimate equation (2) for all states that passed Three Strikes Laws, including California, and for all such states excluding California, in order to make sure the estimated effects of TSL on pleading guilty are not simply picking up effects in California.

(3.2) $$Plea_{ist} = \alpha_0 + \alpha_1 After_t + \alpha_2 Treat_i + \alpha_3 After_t * Treat_i + \gamma X_i + \theta_s + \tau_t + \varphi_j + \pi_{st} + \varepsilon_{ijt}$$

The marginal effects reported in Table 5 can be summarized as follows: I find heterogeneous effects depending on whether the defendant has prior violent convictions. Potential first strikers do not adjust their probability of pleading guilty relative to those who are arrested for a non –violent crime, in response of the implementation of Three Strikes Laws. On the other hand, defendants with at least one prior violent conviction or strike are between 4 to 6 percentage points less likely to plead guilty relative to those with a current non –violent arrest, in response to the implementation of TSL. This effect is statistically significant and robust to the inclusion of state and year interacted, demographics, characteristics of current arrest, and criminal history. The standard errors are clustered by state. Furthermore, these effects are larger when I restrict the analysis to TSL states excluding California. [93]

3.4.3. All States that Did Not Pass Three Strikes Laws

Table 6 estimates equation (2), restricting the universe to the states that never passed TSL. I use an artificial "after" which is an indicator with the value of 1 if the year is after 1994.

[92] Because the California Supreme Court expanded the number of juvenile adjudications that can count as a strike (People v. Davis 1997), I include minors in my analysis of California (Packel, 2002). For the remainder of the paper, I focus on cases filed when the defendant was older than 18.

[93] In California, those who already have one prior violent conviction (one strike), can become a second or third striker with a felony regardless of whether it is violent or non-violent. Because the treatment group is composed of those who are arrested for a violent felony and the control group is composed of those arrested for a non-violent felony, both groups should respond similarly to the implementation of TSL in California, past the first strike. In TSL states other than California, not only the first strike but also the second and third strikes are convictions for violent offenses. This is why the effect is larger at higher strikes when I exclude California.

Because the survey is biannual, and most TSL passed in 1994 or 1995, the variable "after" has a value of 1 for years after 1996 for all TSL states except for Washington.

Table 6 reports that the defendants arrested for a violent offense do not respond differently that those arrested for non-violent felonies in adjusting their probability of pleading guilty in response of this artificial TSL implementation. This result remains insignificant regardless of the specification and incorporation of variables. The standard errors are clustered by state. This provides evidence that the effects described in TSL states are not driven by defendants arrested for violent offenses changing their behavior for another reason that occurred at the same time as TSL was implemented.

3.4.4. Difference in Difference in Difference

As a robustness check, I estimate a difference in difference in difference (DDD) as specified in the following equation

(3.3) $Plea_{ist} = \alpha_0 + DDD_{ist} + \gamma X_i + \theta_s + \tau_t + \varphi_j + \pi_{st} + \varepsilon_{ijt}$

Where DDD_{ist} is defined as follows

(3.4) $DDD_{ist} = \varphi_1 After_{ist} + \varphi_2 Treat_{ist} + \varphi_3 State(TSL)_{is} + \varphi_4 After_{ist} Treat_{ist} + \varphi_5 After_{ist} State(TSL)_{is} + \varphi_6 Treat_{ist} State(TSL)_{is} + \varphi_7 After_{ist} Treat_{ist} State(TSL)_{is}$

I specified $After_{ist}$ to be 1 for years after 1994 (1996 and after) for states that never incorporated TSL. The variable $Treat_{ist}$ has the value of 1 if defendant i is arrested in state s and year t for a violent offense. Finally, $State(TSL)_{is}$ indicates whether the state eventually incorporated TSL

The coefficient φ_7 is the relevant coefficient, and indicates whether the treatment group (arrested for violent felony or potential striker) changes relative to the control group (arrested for non-violent offenses) in response to the implementation of TSL, in the TSL states. Table 7 reports φ_7 for all defendants (Panel A), then separately for defendants with at least one prior violent conviction, or potential second/third strikers (Panel B), then separately for defendants with no prior violent convictions, or potential first strikers (Panel C).

In summary, TSL decreases the probability of pleading guilty between 4 to 6 percentage points among those with at least one prior violent conviction.

3.4.5. Does the probability of pleading guilty to a lesser charge change in response to TSL?

The regression reported in this section is identical to equation 2, with two exceptions. First, the universe now comprises all cases that resulted in conviction and were resolved through plea bargaining. Second, the dependent variable is now an indicator of whether the conviction was a guilty plea to a misdemeanor, a felony, or a violent felony[94] . I apply the same restrictions as before, where I restricted the study to non-dismissed cases where the defendant was at least 18 years old when the case was filed.

Table 8 reports that TSL states, excluding California, do not have a statistically significant change in the distribution of guilty pleas that resolved misdemeanor, felony or even violent felony charges. This is evidence that the decrease in the number of cases obtained by

[94] Felony or misdemeanor are mutually exclusive and hence a 10 percentage point increase in one category implies a decrease of the same magnitude in the other category

guilty pleas are not driven by an increase in the number of prosecutors offering plea bargains with harsher terms, but defendants having a higher propensity to plead guilty.

On the other hand, California seems to have a decrease in the share of cases that were resolved via pleading guilty to violent charges, while there was an increase in the cases that were resolved via pleading guilty to a felony. It is unclear whether this result is driven by prosecutors being more likely to offer guilty pleas for potential strikers in the case of non-violent felonies, or potential strikers being less likely to accept a guilty plea for a violent offense because now they would prefer to go to trial than to accept a guilty plea to a violent offense. My study focuses on the probability of pleading guilty regardless of whether the defendant pleaded guilty to a lesser charge.

3.5. Results

The results show that the introduction of Three Strikes laws significantly reduce the number of criminal cases that are settle with a plea bargain, imposing a potentially costly burden on the legal system.

This occurred even in cases where the passage of Three Strikes legislation did not change the sentence length for the current conviction, but only the future path if the defendant was to commit a crime in the future (i.e. the first strike in California). A simple dynamic model suggests that these laws decrease the continuation value of having a conviction for an offense in the strike-zone. Therefore, I expect an increase in the number of criminal cases that go to trial, rather than being settled with a plea bargain, since the threat of higher future sentences increases the cost of a being convicted for a strikeable offense.

My study finds that, in California, defendants facing the threat of conviction for a first strike are 3 to 4 percentage points less likely to plead guilty after the passage of Three Strikes Laws. Because the length of incarceration for the first violent conviction remains unaffected after the implementation of TSL, this is evidence that defendants are forward-looking and take their future criminal career path into account when making decisions. I could not test what happens at the second or third strike because the two-strike law in California increases sentence length for defendants who are arrested for a felony (violent or non-violent) if they have a previous violent conviction, and the control group is composed of defendants arrested for non-violent felonies.

Excluding California, defendants facing the threat of conviction for a second/third strike are 8 to 9 percentage points less likely to plead guilty following TSL than defendants who are not facing the same threat. These effects are robust to the inclusion of several characteristics, and to a difference-in-difference-in-difference (DDD) estimation, indicating that a potential striker in TSL states is less likely to plead guilty than a "potential striker" in a state where TSL laws were never adopted. This effect is statistically significant among those with at least one prior violent conviction.

This study provides evidence that Three Strikes Laws have the unintended consequence of decreasing the number of filed cases that are resolved by guilty pleas, imposing a costly burden of trials on the legal system. This work is, to the best of my knowledge, the first to provide empirical evidence that defendants are forward-looking and react to the three strikes law by fighting the arrest charge by going to trial instead of pleading guilty, in order to avoid being convicted of a crime that is in the strike category.

Because of the cost of having these additional cases resolved by trial, my study shows the importance of accounting for this type of pre-trial decision when estimating the total cost of Three Strikes.

Bibliography

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

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

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.

Austin, James (1994). `Three strikes and you're out': The likely consequences on the courts, prisons, and crime in California and Washington State, St. Louis University Public Law Review XIV (1).

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, 675-700

Bjerk, David (2005). "Making The Crime Fit the Penalty: The Role of Prosecutorial Discretion Under Mandatory Minimum Sentencing." Journal of Law and Economics, vol. XLVIII

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

Bureau of Justice Assistance (1996) National Assessment of Structured Sentencing. Washington, D.C. U.S. Department of Justice,

Cameron, A. Colin and Pravin Trivedi. (2005) Microeconometrics: Methods and Applications. Cambridge, University Press. New York: NY.

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

Carpenter, Christopher and Carlos Dobkin. (2009) "The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age"; American Economic Journal: Applied Economics 1(1): 164-182.

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

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

Carpenter, Christopher. (2005), Heavy Alcohol Use and the Commission of Nuisance Crime: Evidence from Underage Drunk Driving Laws. *American Economic Review Papers and Proceedings* 95(2): 267-272.

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

Carpenter, Christopher, Lloyd Johnston, Deborah D. Kloska and Patrick O'Malley. (2007) Alcohol Control Policies and Youth Alcohol Consumption: Evidence from 28 Years of Monitoring the Future. B. E. Journal of Economic Analysis and Policy: Topics in Economic Analysis and Policy, 7(1): 1-21.

Carpenter, Christopher and Carlos Dobkin (2010) The Drinking Age, Alcohol Consumption and Crime. Unpublished Manuscript. University of California at Irvine.

Cawley, John, (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, Frank, Rosalie Pacula, Matthew Farrelly, Lloyd Johnston, and Patrick. O'Malley, and Jeremy Bray (1999), Do higher cigarette prices encourage youth to use marijuana? National Bureau of Economic Research, Working Paper No. 6939

Chaloupka, Frank and Weschler, Henry. (1996) Binge Drinking in College: The Impact of Price, Availability and Alcohol Control Policies. Contemporary Economic Policy; Vol. 14, Issue 4: 112-124.

Chaloupka, Frank, Rosalie Liccardo Pacula, Henry Wechsler, Jenny Williams. (2004). Alcohol and Marijuana Use Among College Students: Economic Complements or Substitutes? Health Economics; 13: 825-843

Chesher, Andrew and Margaret Irish, (1987), Residual analysis in the grouped and censored normal linear model. Journal of Econometrics 34, 33-61.

Clark, John, James Austin, and Henry D. Alan (1997) Three Strikes and You're Out: A Review of State Legislation. Washington D.C. National Institute of Justice.

Countywide Criminal Justice Coordination Committee (1995), Impact of the `Three Strikes Law' on the Criminal Justice System in Los Angeles County

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

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

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

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.

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

Grossman, Gene and Michael, Katz (1983) “Plea Bargaining and Social Welfare.” 73 American Economic Review 749-57

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 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.

Imbens, Guido and Thomas Lemieux. “Regression Discontinuity Designs: A Guide to Practive.” Journal of Econometrics (2008) 142: 615-635.

Iyengar, Radha (2008). “I’d rather be Hanged for a Sheep than a Lamb: The Unintended Consequences of Three Strikes’ Laws.” NBER Working Paper No. 13784.

Joksch, Hans and Ralph Jones. “Changes in the Drinking Age and Crime.” Journal of Criminal Justice (1993) 21(3): 209-221.

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, Donald, Alan Mathios and Rosalie L. Pacula. “Economics of Youth Drug Use, Addiction and Gateway Effects.” Addiction 96(2001): 151-164.

Kuziemko, Ilyana (2006). “Does the Threat of the Death Penalty Affect Plea Bargaining in Murder Cases? Evidence from New York’s 1995 Reinstatement of Capital Punishment.” American Law and Economics Review.

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. “Alcohol, drugs and violent crime.” International Review of Law and Economics (2005) 25:20-44.

Marvell, Thomas B. and Moody, Carlisle (2001). “The Lethal Effects of Three Strikes Laws.” Journal of Legal Studies 30; 89-106

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.

Michael G. Turner et al. (1995) 'Three Strikes and You're Out' Legislation: A National Assessment," Federal Probation, Vol. 59, No. 8.

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.

Oren, Bar-Gill, and Gazal Oren, (2004) Plea Bargains Only for the Guilty. Harvard Law School John M. Olin Center for Law, Economics and Business Discussion Paper Series, Paper 481. http://lsr.nellco.org/harvard_olin/481.

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

Packel, Amanda (2002) Juvenile Justice and the Punishment of Recidivists Under California's Three Strikes. California Law Review Volume:90 Issue 4; 1157 to 1202

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

Pacula, Rosalie.(1997)Economic Modeling of the Gateway Effect." Health Economics 6(5): 521-524

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.

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.

Shepherd, Joanna (2002). “Fear of the First Strike: The Full Deterrent Effect of California’s Two and Three Strike Legislation.” Journal of Legal Studies, vol XXXI.

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.

Tonry, Michael (1996) Sentencing Matters. New York: Oxford University Press.

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

U.S. Dept. of Justice, Bureau of Justice Statistics (2010) State Court Processing Statistics, 1990-2006: Felony Defendants in Large Urban Counties [Computer file]. Conducted by Pretrial Justice Institute (formerly, the Pretrial Services Resource Center)[producer]. ICPSR02038-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

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.

Walsh, Jennifer Edwards (1999) `In the Furtherance of Justice’: The Effect of Discretion on the Implementation of California’s Three Strikes Law. Paper presented at the American Political Science Association annual meeting, Atlanta, September2-5.

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.

Figures and Tables

Table 1.1: Summary Statistics

		Non-missing Drug Questions for all 10 waves			
			Ever Used		
	Full Sample		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. 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.
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 1.2: Parameter Estimates for the Models with Homogeneous State Dependence and Stepping-Stone

	Model A: Logit			Model B: Probit			Model C: Logit		
	Alc.	**Mar.**	**Coc.**	**Alc.**	**Mar.**	**Coc.**	**Alc.**	**Mar.**	**Coc.**
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)
$\mathbf{cov}(\varepsilon_{it}^{drink}, \varepsilon_{it}^{mar})$					0.58				
$\mathbf{cov}(\varepsilon_{it}^{drink}, \varepsilon_{it}^{coc})$					0.38				
$\mathbf{cov}(\varepsilon_{it}^{mar}, \varepsilon_{it}^{coc})$					0.46				
$Y_{j,t-2}$							0.99	0.68	1.01
							(0.05)	(0.06)	(0.09)
$Y_{j,t-1} * Y_{j,t-2}$							0.13	-0.04	-0.56
							(0.06)	(0.08)	(0.12)
N		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. These parameters are the logit or probit estimates, not the marginal effects.
Model A: Trivariate logit model with first-order state-dependence
Model B: Trivariate probit model with first-order state-dependence
Model C: Trivariate logit model with second-order state-dependence

Table 1.3 : Marginal Effects

	Model A	Model B	Model C		Model D		
			Y_{t-2}=0	Y_{t-2}=1	Age 25	Age 20	Age 15
Panel A: State Dependence							
Alcohol	22.59	25.56	34.18	20.53	29.97	26.16	23.69
Marijuana	15.29	17.65	13.58	25.80	21.28	19.99	13.79
Cocaine	6.95	10.14	6.11	25.20	7.23	8.61	5.43
Panel B: Stepping Stone Effect							
From alcohol to mar.	4.62	10.13	4.14		-4.24	0.71	6.42
From alcohol to cocaine	1.45	3.49	0.84		-1.06	0.44	2.21
From mar. to cocaine	1.32	2.50	1.36		1.21	2.76	3.24

Note: The marginal effects are computed using simulated data because a random effect needs to be assigned. They are averaged over the years.

Table 1.4: 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.. For instance, the measure of alcohol persistence in the column Sim($\gamma_{kj} = 0$) represents what the measure of alcohol persistence would be in the absence of alcohol state-dependence.

Table 1.5: 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

Table 1.5: (Continued)

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 1.6: 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)	2000	34.32	31.52	31.80	30.80	34.32	32.92
(1,0,0)	2000	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

Table 1.6 (Continued)

(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)	2003	25.80	23.02	21.50	20.10	25.80	24.15
(1,0,0)	2003	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)	2004	22.28	21.87	19.20	19.30	22.28	22.71
(1,0,0)	2004	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

Table 1.6 (Continued)

(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.

Table 1.7: Sample Analogue of Generalized Residuals

Panel A: Mean, Variance, and Serial Correlation of Generalized Residuals

	Model A			Model B			Model C		
	drink	**mar**	**coc**	**drink**	**mar**	**coc**	**drink**	**mar**	**coc**
Mean	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)
Variance	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)
First Order	-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)
Second Order	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)
Third Order	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)
Fourth Order	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)
Fifth Order	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)

Table 1.7 (Continued)

Panel B: Correlation of Generalized Residuals Across Drugs and Time

	Model A			Model B			Model C		
	coc-drink	**mar-drink**	**coc-mar**	**coc-drink**	**mar-drink**	**coc-mar**	**coc-drink**	**mar-drink**	**coc-mar**
Correlation	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)
First Order	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)
Second Order	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)
Third Order	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)
Fourth Order	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)
Fifth Order	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 1^{st}-5^{th} order autocorrelation of sample-analogue generalized residuals for alcohol marijuana, and cocaine. Panel B corresponds to the sample-analogue of generalized residuals across drugs.

Table1.8: 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		
	Alcohol	**Mar**	**Coc**	**Alcohol**	**Mar**	**Coc**	**Alcohol**	**Mar**	**Coc**	**Alcohol**	**Mar**	**Coc**
Constant Part of The Stepping-Stone Effect												
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
	(0.20)	(0.26)	(0.44)	(0.04)	(0.07)	(0.12)	(0.08)	(0.09)	(0.13)	(0.08)	(0.09)	(0.13)
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
	(0.31)	(0.25)	(0.37)	(0.07)	(0.05)	(0.08)	(0.10)	(0.06)	(0.07)	(0.10)	(0.06)	(0.07)
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
	(0.56)	(0.41)	(0.40)	(0.12)	(0.08)	(0.08)	(0.14)	(0.07)	(0.06)	(0.14)	(0.07)	(0.06)
Heterogeneour Part of the Stepping-Stone Effect												
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
	(0.01)	(0.01)	(0.02)	(0.06)	(0.09)	(0.16)	(0.05)	(0.07)	(0.14)	(0.05)	(0.12)	(0.22)
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
	(0.02)	(0.01)	(0.02)	(0.09)	(0.07)	(0.11)	(0.17)	(0.05)	(0.11)	(0.07)	(0.05)	(0.09)
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
	(0.03)	(0.02)	(0.02)	(0.16)	(0.11)	(0.11)	(0.20)	(0.09)	(0.15)	(0.13)	(0.10)	(0.15)
LL	-48650				-48792			-48766			-48765	
Nparams	68				68			68			68	
AIC	97435				97720			97668			97666	
BIC	97880				98164			98113			98111	
N	5108				5108			5108			5108	

Table 1.8 (Continued)

	Model H: by Intensity		
	Alcohol	**Mar**	**Coc**
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		-69395	
Nparams		74	
AIC		138937	
BIC		139417	
N		4847	

Table 1.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 1.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 1.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 2.1: Summary Statistics

	Full sample	Subsample of respondents not lost due to attrition				
			Non-missing drug-related questions			
			Alcohol	**Marijuana**	**Cigarette**	**Cocaine**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Demographics						
Age(1997)	14.35	14.24	14.24	14.24	14.24	14.24
	(1.49)	(1.47)	(1.47)	(1.47)	(1.47)	(1.47)
Male	0.51	0.47	0.46	0.46	0.47	0.46
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Black	0.26	0.27	0.25	0.25	0.26	0.26
	(0.44)	(0.44)	(0.43)	(0.43)	(0.44)	(0.44)
Hispanic	0.21	0.20	0.21	0.21	0.20	0.20
	(0.41)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)
Mixed race	0.01	0.01	0.01	0.01	0.01	0.01
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
White	0.52	0.52	0.53	0.53	0.53	0.53
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
P(ever alcohol)	0.93	0.95	0.95	0.95	0.95	0.95
	(0.25)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
P(ever marijuana)	0.59	0.60	0.59	0.58	0.59	0.59
	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)
P(ever cocaine)	0.24	0.26	0.26	0.25	0.26	0.25
	(0.43)	(0.44)	(0.44)	(0.43)	(0.44)	(0.43)
P(ever cigarette)	0.73	0.73	0.72	0.71	0.72	0.72
	(0.44)	(0.45)	(0.45)	(0.45)	(0.45)	(0.45)
Panel B: Average Starting Age			**Among those who Ever Used**			
	All		**Alcohol**	**Tobacco**	**Mar**	**Coc**
Alcohol Starting Age	15.42	15.45	15.45	14.73	14.41	14.01
Tobacco Starting Age	15.04	15.13	15.11	15.13	14.59	14.01
Marijuana Starting Age	16.61	16.79	16.79	16.52	16.79	15.87
Cocaine Starting Age	18.27	18.34	18.35	18.30	18.24	18.34
Percentage	100.00	59.94				
N	8984	5385				

Column (1) : Entire Sample

Column (2):Sample consists of respondents who answer the questionnaire from 1997 to 2009

Columns (3)-(6): Subsample of column (2) who consumed in at least one period alcohol, marijuana, tobacco and cocaine respectively.

Table 2.2: Age Profile of Demographic Characteristics

	Attending College		Male		Black		Hispanic		White	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Over 21=Di	-0.006	0.008	0.002	-0.002	-0.002	-0.014	0.009	0.004	-0.011	0.005
	(0.008)	(0.013)	(0.009)	(0.014)	(0.008)	(0.012)	(0.007)	(0.011)	(0.009)	(0.014)
Constant	0.224***	0.217***	0.533***	0.533***	0.329***	0.331***	0.182***	0.186***	0.484***	0.478***
	(0.039)	(0.039)	(0.047)	(0.047)	(0.044)	(0.044)	(0.037)	(0.037)	(0.047)	(0.047)
cluster (id)	Yes		Yes		Yes		Yes		Yes	
Year Effects	Yes		Yes		Yes		Yes		Yes	
N clusters	8542									
N	31879									

Note:

Model 1: Model with first order interacted polynomial

Model 2: Model with second-order interacted polynomial .

The reported coefficients are estimated by a OLS model with year effects ,and SE are clustered at the individual level.

Table 2.3: Measures of Alcohol Participation

	Y1=Alcohol Last Year			Y2=Alcohol Last Month			Y3=Binge Drink Last month		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Over 21=Di	0.056***	0.058***	0.058**	0.079***	0.078***	0.083***	0.059***	0.062***	0.068**
	(0.008)	(0.012)	(0.017)	(0.009)	(0.013)	(0.020)	(0.009)	(0.013)	(0.020)
cluster (id)	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
N clusters		8526			8524			8542	
N		31591			31483			31879	

	Y4=% Last Month Drink			Y5=% Last Month Binge Drink		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Over 21=Di	0.039***	0.037***	0.031***	0.013***	0.015***	0.014**
	(0.004)	(0.005)	(0.008)	(0.002)	(0.004)	(0.005)
cluster (id)	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y
N clusters		8523			8522	
N		31416			31273	

Note: Outcome 4 (5) is the number of days in the last month where the respondent reported drinking (binged drinking) divided by 30

Model 1,2 and 3 include a first, second, and third order age-centered interacted polynomial respectively.

These coefficients were estimated using a panel fixed effects regression model with year effects, individual level FE, and SE clustered at the individual level.

Table 2.4: Measures of Cocaine Participation and Criminal Activity (Measures of Last Year)

Panel A: Measures of Cocaine Consumption

	Y1=Cocaine Last Year			**Y2=Percentage of Days in Last Year in which Cocaine was Consumed**								
				All (a)			Users Last Year(b)			Ever Users(c)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Over 21=Di	-0.014**	-0.018**	-0.017	0.000	0.001	0.001	0.036	0.093*	0.112*	0.005	-0.003	0.003
	(0.005)	(0.007)	(0.011)	(0.003)	(0.004)	(0.006)	(0.024)	(0.038)	(0.056)	(0.007)	(0.011)	(0.015)
cluster (id)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N clusters		8475			8542			1185			2129	
N		31345			31879			1952			8078	

Panel B: Measures of Criminal Activity

	Y1=Stealing			**Y2=Drug Dealing**			**Y4=Fight Participation**			**Y4=Property Destruction**		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Over 21=Di	-0.015**	-0.029**	-0.029*	-0.006	0.003	-0.006	0.001	-0.002	0.001	-0.004	-0.007	-0.017
	(0.006)	(0.009)	(0.013)	(0.005)	(0.008)	(0.012)	(0.006)	(0.009)	(0.014)	(0.005)	(0.007)	(0.011)
cluster (id)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N clusters		8396			8392			8395			8396	
N		24677			24650			24666			24653	

Panel C: Measures of Initiation to Other Drugs

	Alcohol			**Tobacco**			**Marijuana**			**Cocaine**		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Over 21=Di	0.015**	0.012	0.016	0.002	0.000	0.000	-0.003	0.000	-0.012	-0.011**	-0.013*	-0.015
	(0.005)	(0.007)	(0.010)	(0.004)	(0.006)	(0.009)	(0.004)	(0.007)	(0.010)	(0.004)	(0.006)	(0.009)
cluster (id)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N clusters		8010			7938			7693			7367	
N		30214			29961			29121			27914	

Note:

Model 1,2 and 3 include a first, second, and third order age-centered interacted polynomial respectively.

Models do not include covariates because demographic variables are time-invariant and would be eliminated with a fixed effects model.

Panel A: Two outcomes (1) Indicator for whether the respondent consumed cocaine in the year prior to the interview, (2) the number of days in the year prior to the interview in which the respondent consumed cocaine divided over 365.

The RD parameters for the second outcome are estimated on three subsamples. (a)Frequency of cocaine consumption last year among all respondents. (b)Frequency of cocaine consumption last year among those who used cocaine in the last year. (c)Frequency of cocaine consumption last year among those who used cocaine at least once between 1997 and 2009 .

Table 2.5: Robustness Checks

	Full Panel		Balanced Panel	
Panel A: Alcohol Consumption				
Alcohol Last Year	0.068***	0.059***	0.079***	0.077***
	(0.009)	(0.012)	(0.011)	(0.014)
Alcohol Last Month	0.095***	0.077***	0.100***	0.090***
	(0.011)	(0.013)	(0.013)	(0.016)
Binge Drinking Last Month	0.067***	0.061***	0.065***	0.080***
	(0.010)	(0.013)	(0.013)	(0.016)
%Days Drink Last Month	0.045***	0.037***	0.042***	0.041***
	(0.004)	(0.005)	(0.005)	(0.007)
%Days Binge Drink Last Month	0.015***	0.013***	0.016***	0.017***
	(0.003)	(0.004)	(0.003)	(0.004)
Panel B: Cocaine Consumption				
Cocaine last year	-0.013*	-0.019**	-0.014*	-0.020*
	(0.006)	(0.007)	(0.007)	(0.009)
%Days Last Year Cocaine	0.002	0.000	0.000	0.005
	(0.003)	(0.005)	(0.003)	(0.005)
%Days Last Year Cocaine(Univ:Users Last Year)	0.021	0.077*	0.037	0.072
	(0.030)	(0.038)	(0.033)	(0.041)
%Days Last Year Cocaine(Univ:Ever Users)	-0.001	-0.005	0.002	0.014
	(0.009)	(0.012)	(0.010)	(0.013)
Panel B: Drug Initiation				
Alcohol	0.013*	0.013	0.026***	0.021*
	(0.005)	(0.007)	(0.007)	(0.008)
Tobacco	0.003	0.000	0.003	0.003
	(0.005)	(0.006)	(0.006)	(0.007)
Marijuana	0.001	0.001	0.006	0.006
	(0.005)	(0.007)	(0.006)	(0.008)
Cocaine	-0.008	-0.013*	-0.006	-0.015*
	(0.005)	(0.006)	(0.006)	(0.007)

Table 2.5: Robustness Checks (Continued)

Panel C: Criminal Participation				
Stealing	-0.025***	-0.030**	-0.024**	-0.026*
	(0.007)	(0.009)	(0.008)	(0.010)
Fighting	0.002	-0.002	-0.001	0.000
	(0.007)	(0.009)	(0.008)	(0.011)
Property Destruction	-0.003	-0.007	0.000	0.005
	(0.006)	(0.008)	(0.007)	(0.009)
Drug Dealing	-0.005	0.004	-0.004	0.006
	(0.006)	(0.008)	(0.008)	(0.010)
Full Panel	X	X		
Balanced Panel			X	X
Age Bandwidth 18-24	X		X	
Age Bandwidth 19-23		X		X
Cluster ID	X	X	X	X
Year Effects	X	X	X	X
Birthday Effects	X	X	X	X

Note: These parameters are estimated with a panel fixed effects regression model with a second order age-centered polynomial. All specifications include year effects, birthday effects, and cluster the standard errors at the individual level.

Table 3.1:Three Strikes Laws Passed Between 1993 and 1995

St.	Features of Three Strikes Legislation	Yr	Features of Preexisting Sentencing Laws
CA	Mandatory doubling of sentence for any felony if one prior serious or violent-felony conviction; mandatory life without parole for 25 years for any third felony conviction if two prior serious or violent-felony convictions.	1994	Life with no parole eligibility before 20 years for third violent -felony conviction where separate prison terms were served for the first two; life with no parole for fourth violent-felony conviction
FL	Added new category of "violent career criminal" to existing habitual offender statute; for third conviction for specified violent offense, life if first-degree felony, 30-40 years if second-degree felony, 10-15 years for third-degree felony.	1995	Categories of habitual felony offender and habitual violent offender; range of enhanced sentences.
GA	Mandatory life without parole for second specified violent felony conviction.	1995	On fourth felony conviction, offender must serve maximum time imposed and not be eligible for parole until maximum sentence served.
IN	Mandatory life without parole for second specified violent felony conviction	1994	Habitual offender law requiring enhanced sentencing on third felony conviction.
MD	Life without parole for fourth violent-felony conviction for which separate prison terms were served for the first three.	1994	Same law, expect that carjacking and armed carjacking were not on the list of offenses receiving this sentence.
NJ	Mandatory life without parole for third conviction for certain violent felonies.	1995	Rarely invoked "persistent offender" provision allowing sentence one degree higher than the conviction offense on third conviction for first-, second-, or third-degree -felonies.

(Continued) Table 3.1

PA	Mandatory minimum enhanced sentence of 10 years for second conviction for crime of violence and 25 years for third such conviction.	1995	Mandatory minimum enhanced sentence of 5 years for second or subsequent conviction for certain specified crimes of violence.
TN	Mandatory life without parole for second conviction for designated violent felonies; same for third conviction for other violent felonies.	1994	Mandatory life without parole for third violent-felony conviction.
UT	Second - and third-degree felony offenders sentenced as first-degree felons and first-degree felons not eligible for probation if they have two prior convictions for any felony and a present conviction for a violent felony.	1995	Second-and third-degree felonies receive enhanced sentence of 5 years to life if offender has two prior convictions at least as severe as second-degree felonies.
VA	Mandatory life without parole on third conviction for specified violent felonies or drug distribution charges.	1994	No parole eligibility if convicted of three separate violent felonies.
WA	Mandatory life without parole on third conviction for specified violent felonies.	1993	Number of prior convictions factored into offender score on state's sentencing guidelines.
WI	Mandatory life without parole on third conviction for specified serious offenses.	1994	For repeat felony offenders, up to 10 years can be added to sentences of 10 years or more; 6 years can be added to sentences of 1-10 years.

Sources: This table is equivalent to Table 1 from Bjerk (2005)

For a discussion of Three Strikes Laws by State, see Clark, John, James Austin, and Alan Henry (1997)

Table 3.2: Summary Statistics

	California		States with TSL		States Without TSL	
	Violent	Non-violent	Violent	Non-violent	Violent	Non-violent
Panel A: All Defendants						
Age	29.66	31.32	30.15	30.88	28.65	30.21
	(10.27)	(9.70)	(10.67)	(9.88)	(10.37)	(10.14)
White	0.27	0.35	0.33	0.38	0.36	0.39
	(0.45)	(0.48)	(0.47)	(0.49)	(0.48)	(0.49)
Black	0.28	0.25	0.47	0.41	0.50	0.48
	(0.45)	(0.43)	(0.50)	(0.49)	(0.50)	(0.50)
Male	0.91	0.81	0.86	0.82	0.88	0.83
	(0.29)	(0.39)	(0.34)	(0.39)	(0.33)	(0.38)
N Previous Convictions	2.57	3.35	2.42	3.11	2.02	2.40
	(3.12)	(3.45)	(3.21)	(3.50)	(2.97)	(3.12)
N Previous Felony Convictions	0.90	1.38	1.07	1.46	0.78	1.04
	(1.62)	(2.10)	(2.01)	(2.32)	(1.59)	(1.80)
N Previous Misd Convictions	1.81	2.28	1.58	2.00	1.37	1.54
	(2.62)	(2.88)	(2.56)	(2.81)	(2.48)	(2.56)
Prob. Prior Conviction	0.61	0.71	0.55	0.65	0.50	0.58
	(0.49)	(0.45)	(0.50)	(0.48)	(0.50)	(0.49)
Prob. Prior Felony Conviction	0.37	0.48	0.37	0.46	0.32	0.41
	(0.48)	(0.50)	(0.48)	(0.50)	(0.47)	(0.49)
Prob. Prior Misd. Conviction	0.53	0.61	0.46	0.54	0.40	0.45
	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)
Prob. Prior Violent Conviction	0.17	0.12	0.17	0.13	0.14	0.10
	(0.37)	(0.32)	(0.38)	(0.34)	(0.35)	(0.30)
Prob. Current Conviction	0.82	0.85	0.62	0.74	0.56	0.69
	(0.38)	(0.36)	(0.49)	(0.44)	(0.50)	(0.46)

Continued (Table 3.2)

Panel B: Conditional on Conviction

Prob. Conviction	1.00	1.00	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Felony Conviction	0.86	0.91	0.81	0.86	0.73	0.77
	(0.34)	(0.28)	(0.39)	(0.34)	(0.45)	(0.42)
Misd. Conviction	0.14	0.09	0.19	0.14	0.27	0.23
	(0.34)	(0.28)	(0.39)	(0.34)	(0.45)	(0.42)
Violent Conviction	0.75	0.00	0.71	0.00	0.64	0.01
	(0.43)	(0.06)	(0.46)	(0.05)	(0.48)	(0.07)
Violent Conviction if Guilty Plea	0.74	0.00	0.70	0.00	0.62	0.01
	(0.44)	(0.06)	(0.46)	(0.05)	(0.48)	(0.07)
Guilty Plea	0.94	0.99	0.90	0.95	0.92	0.96
	(0.23)	(0.11)	(0.29)	(0.22)	(0.28)	(0.21)
N	5134	18928	12931	39570	10675	34039

Notes: States with TSL are states that eventually passed Three Strikes Laws (TSL)

This Table reports summary statistics for respondents who are currently arrested for a violent and non-violent offense, separetely.

SD are in parenthesis

Universe: Non-pending cases with non-missing adjudication outcome, non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race).

Table 3.3: Probability of Pleading Guilty (States with Three Strikes Laws)

Panel A: Compare Before and After in TSL States

	Violent Arrest		**Non-Violent Arrest**		**Difference**
	Pre-TSL	**Post-TSL**	**Pre-TSL**	**Post-TSL**	
Prob. Plea	0.857	0.830	0.880	0.871	-0.016
SD	(0.35)	(0.38)	(0.32)	(0.34)	(0.010)
N	2126	6324	7785	23986	

Panel B: California

	Violent Arrest		**Non-Violent Arrest**		**Difference**
	Pre-TSL	**Post-TSL**	**Pre-TSL**	**Post-TSL**	
Prob. Plea	0.943	0.926	0.926	0.928	-0.030*
SD	(0.23)	(0.26)	(0.26)	(0.26)	(0.013)
N	1054	3168	4134	12970	

The sample includes all defendants older than 18 in non-TSL states with non-missing or pending adjudication outcome, , non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). Non-dismissed Cases only. Coefficients (Difference) represent the marginal change in probability (from a probit model). Statistics weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties.

SD are in parenthesis

(*) significant at the 5* level

(**)significant at the 1% level

(***)significant at the 0.1% level

Table 3.4: Probability of Pleading Guilty Versus Going to Trial in California (Defendants Older than 18)

	All Ages					Adults (At least Age 18)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Panel A: All Defendants Regardless of their Prior Convictions										
Treat X After	-0.027*	-0.030**	-0.0326**	-0.047***	-0.0372***	-0.030*	-0.033**	-0.033**	-0.046***	-0.035**
	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Log likelihood	-5609.13	-5450.15	-5431.12	-4786.28	-4152.31	-5538.15	-5377.28	-5376.44	-4742.13	-4117.52
N	21416	21416	21416	21408	18939	21326	21326	21326	21318	18866
Panel B: Defendants with At Least One Prior Violent Conviction										
Treat X After	-0.035	-0.032	-0.031	-0.041	-0.028	-0.035	-0.032	-0.031	-0.041	-0.028
	(0.037)	(0.036)	(0.034)	(0.037)	(0.038)	(0.037)	(0.036)	(0.034)	(0.037)	(0.038)
Log likelihood	-463.08	-453.25	-446.70	-415.96	-393.96	-462.97	-453.15	-446.70	-415.96	-393.96
N	1983	1983	1983	1939	1817	1982	1982	1982	1939	1817
Panel C: Defendants with No Prior Violent Convictions										
Treat X After	-0.017	-0.020	-0.027	-0.043**	-0.039**	-0.023	-0.026	-0.026	-0.041**	-0.036**
	(0.016)	(0.015)	(0.016)	(0.017)	0.0160	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)
Log likelihood	-3885.23	-3738.27	-3719.12	-3220.82	-2994.38	-3822.57	-3675.14	-3673.56	-3186.47	-2960.32
N	14138	14138	14138	14130	13623	14065	14065	14065	14057	13551
Year effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
Demographics (a)	N	N	Y	Y	Y	N	N	Y	Y	Y
Current arrest (b)	N	N	N	Y	Y	N	N	N	Y	Y
Criminal history (c)	N	N	N	N	Y	N	N	N	N	Y

Notes: The sample includes all defendants in California with non-missing or pending adjudication outcome, , non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). Non-dismissed cases only.

(a) Demographics:male, black, age. (b) Current arrest:total number of current arrest charges, and an indicator for the most serious arrest charge (list of 16 categories)

(c)Criminal history: Dummy for whether the defendant has any prior (and if so, how many) convictions, felony convictions, misdemeanor convictions, and incarcerations.

Coefficients represent the marginal change in probability (from a probit model). Statistics are weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties. SD are in parenthesis. (*) significant at the 5* level, (**)significant at the 1% level, (***)significant at the 0.1% level.

Table 3.5: Probability of Pleading Guilty Versus Going to Trial (Defendants Older than 18 in States with Three Strike Laws)

	States With Three Strikes Laws					States With Three Strikes Laws (Except CA)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Panel A: All Defendants										
Treat X After	-0.013	-0.011	-0.013	-0.019	-0.008	0.006	0.016	0.006	0.005	0.022
	(0.017)	(0.018)	(0.017)	(0.021)	(0.021)	(0.024)	(0.014)	(0.024)	(0.024)	(0.021)
Log likelihood	-13989.6	-13456.7	-13974.0	-13521.1	-11875.2	-8587.5	-8111.9	-8567.0	-8486.7	-7486.3
N	40221	39870	40221	40205	35903	18895	18544	18895	18887	17037
Panel B: Defendants with At Least One Prior Violent Conviction										
Treat X After	-0.063***	-0.042**	-0.067***	-0.065***	**-0.037**	-0.085*	-0.049	-0.089**	-0.093**	-0.051
	(0.021)	(0.017)	(0.022)	(0.022)	**(0.023)**	(0.037)	(0.033)	(0.038)	(0.038)	(0.035)
Log likelihood	-1524.0	-1440.0	-1519.8	-1484.2	-1337.5	-1062.4	-996.1	-1057.3	-1037.0	-913.3
N	4462	4335	4462	4461	4112	2480	2353	2480	2479	2254
Panel B: Defendants with No Prior Violent Convictions										
Treat X After	-0.007	-0.010	-0.007	-0.014	-0.010	0.013	0.010	0.013	0.015	0.019
	(0.017)	(0.018)	(0.017)	(0.023)	(0.022)	(0.025)	(0.024)	(0.025)	(0.025)	(0.024)
Log likelihood	-9930.1	-9546.6	-9908.8	-9560.2	-9023.2	-6242.3	-5904.4	-6218.3	-6151.7	-5826.4
N	27461	27187	27461	27448	26537	13396	13122	13396	13391	12986
State Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Demographics (a)	N	N	Y	Y	Y	N	N	Y	Y	Y
Current arrest(b)	N	N	N	Y	Y	N	N	N	Y	Y
Criminal history (c)	N	N	N	N	Y	N	N	N	N	Y
State * Year	N	Y	N	N	N	N	Y	N	N	N
cluster by state	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The sample includes all defendants older than 18 in TSL states with non-missing or pending adjudication outcome, , non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). (a) Demographics:male, black, age. (b) Current arrest:total number of current arrest charges, and an indicator for the most serious arrest charge (list of 16 categories) (c)Criminal history: Dummy for whether the defendant has any prior (and if so, how many) convictions, felony convictions, misdemeanor convictions, and incarcerations. Coefficients represent the marginal change in probability (from a probit model). Statistics are weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties. SD are in parenthesis. (*) significant at the 5* level (**)significant at the 1% level, (***)significant at the 0.1% level

Table 3.6: Probability of Pleading Guilty (States without Three Strike Laws)

	Artificial Three Strikes in 1994				
	(1)	(2)	(3)	(4)	(5)
Panel A: All Defendants					
Treat X After	-0.005	0.006	-0.005	-0.001	0.009
	(0.019)	(0.018)	(0.018)	(0.017)	(0.018)
Log likelihood	-9167.4	-8811.2	-9155.1	-9045.5	-7845.9
N	29567	29297	29567	28978	24982
Panel B: Defendants with At Least One Prior Violent Conviction					
Treat X After	-0.030	-0.040	-0.032	-0.023	-0.033
	(0.030)	(0.029)	(0.030)	(0.030)	(0.040)
Log likelihood	-797.4	-749.3	-795.5	-771.1	-669.9
N	2796	2658	2796	2695	2458
Panel B: Defendants with No Prior Violent Convictions					
Treat X After	-0.005	0.007	-0.004	-0.001	0.010
	(0.025)	(0.024)	(0.024)	(0.023)	(0.020)
Log likelihood	-6616.2	-6351.0	-6603.2	-6523.5	-6117.9
N	20424	20213	20424	19945	18890
State Effects	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y
Demographics (a)	N	N	Y	Y	Y
Current arrest(b)	N	N	N	Y	Y
Criminal history (c)	N	N	N	N	Y
State * Year	N	Y	N	N	N
cluster by state	Y	Y	Y	Y	Y

Notes: The sample includes all defendants older than 18 in non-TSL states with non-missing or pending adjudication outcome, non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). Non-dismissed cases only. (a) Demographics:male, black, age. (b) Current arrest:total number of current arrest charges , and an indicator for the most serious arrest charge (list of 16 categories) (c)Criminal history: Dummy for whether the defendant has any prior (and if so, how many) convictions, felony convictions, misdemeanor convictions, and incarcerations. Coefficients represent the marginal change in probability (from a probit model). Statistics are weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties. SD are in parenthesis

(*) significant at the 5* level

(**)significant at the 1% level

(***)significant at the 0.1% level

Table 3.7: Probability of Pleading Guilty (DDD Estimates)

	(1)	(2)	(3)	(4)	(5)
Panel A: All Defendants					
DDD	-0.011	-0.011	-0.010	-0.013	-0.004
	(0.015)	(0.017)	(0.015)	(0.018)	(0.018)
Log likelihood	-23244.8	-22261.0	-23223.6	-22793.7	-19943.7
N	69788	69167	69788	69183	60885
Panel B: Defendants with At Least One Prior Violent Conviction					
DDD	-0.057***	-0.038**	-0.060***	-0.060***	-0.036
	(0.019)	(0.016)	(0.020)	(0.021)	(0.022)
Log likelihood	-2337.4	-2188.1	-2333.9	-2285.1	-2040.6
N	7258	6993	7258	7156	6570
Panel B: Defendants with No Prior Violent Convictions					
DDD	-0.005	-0.009	-0.005	-0.007	-0.005
	(0.015)	(0.016)	(0.015)	(0.018)	(0.018)
Log likelihood	-16615.0	-15894.8	-16586.6	-16260.9	-15328.0
N	47885	47400	47885	47393	45427
State Effects	Y	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y
Demographics (a)	N	N	Y	Y	Y
Current arrest(b)	N	N	N	Y	Y
Criminal history (c)	N	N	N	N	Y
State * Year	N	Y	N	N	N
cluster by state	Y	Y	Y	Y	Y

Note: DDD=Treat*After*Three_strikes_state

Treat indicates whether the defendant is currently arrested for a violent offense

After is defined as 1994 for the states that never adopted the Three Strikes Laws

Three Strikes State indicates whether the respondent is in a state with Three Strikes Law

The sample includes all defendants older than 18 in non-TSL states with non-missing or pending adjudication outcome, , non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). Non-dismissed cases only. (a) Demographics: male, black, age. (b) Current arrest:total number of current arrest charges, and an indicator for the most serious arrest charge (list of 16 categories). (c)Criminal history: Dummy for whether the defendant has any prior (and if so, how many) convictions, felony convictions, misdemeanor convictions, and incarcerations. Coefficients represent the marginal change in probability (from a probit model). Statistics are weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties. SD are in parenthesis. (*) significant at the 5* level. (**)significant at the 1% level, (***)significant at the 0.1% level

Table3. 8: Probability of Pleading Guilty to Misd. or Violent Felony, Conditional on Pleading Guilty

	CA		TSL States (no CA)		TSL States	
	Misd	Violent	Misd	Violent	Misd	Violent
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Defendants						
Treat X After	-0.017	-0.014**	-0.023	0.003	-0.020	-0.005
	(0.010)	(0.004)	(0.030)	(0.007)	(0.010)	(0.004)
Log likelihood	-4899.5	-2203.5	-5788.8	-1812.6	-10711.6	-4068.2
N	17482	16221	13410	12247	30892	30314
Panel B: Defendants with At Least One Prior Violent Conviction						
Treat X After	-0.024	-0.713(d)	0.005	-0.054	-0.016	-0.085
	(0.021)	(0.261)	(0.053)	(0.044)	(0.017)	(0.028)
Log likelihood	-364.7	-243.1	-644.9	-267.3	-1014.8	-520.3
N	1719	850	1804	1054	3535	2374
Panel B: Defendants with No Prior Violent Convictions						
Treat X After	-0.020	-0.011**	-0.028	0.007	-0.025*	-0.003
	(0.011)	(0.005)	(0.030)	(0.010)	(0.010)	(0.005)
Log likelihood	-3412.8	-1467.4	-4429.1	-1298.2	-7863.1	-2804.4
N	12503	11557	10084	8652	22587	21105
State Effects	N	N	Y	Y	Y	Y
Year effects	Y	Y	Y	Y	Y	Y
State*Year	N	N	Y	Y	Y	Y
Demographics (a)	Y	Y	Y	Y	Y	Y
Current arrest(b)	Y	Y	Y	Y	Y	Y
Criminal history (c)	Y	Y	Y	Y	Y	Y
cluster by state	N	N	Y	Y	Y	Y

Universe: Convicted defendants whose conviction was resolved through a guilty plea. The sample includes all defendants older than 18 with non-missing or pending adjudication outcome, non-missing most serious adjudication and conviction charge, non-missing adjudication outcome, and non-missing demographics (age, sex and race). Non-dismissed cases only.

(a) Demographics:male, black, age.

(b) Current arrest:total number of current arrest charges, and an indicator for the most serious arrest charge (list of 16 categories)

(c)Criminal history: Dummy for whether the defendant has any prior (and if so, how many) convictions, felony convictions, misdemeanor convictions, and incarcerations. Coefficients represent the marginal change in probability (from a probit model). Statistics are weighted using the weights provided by State Court Processing Statistics to be representative of the nation's 75 most populous counties. SD are in parenthesis. (*) significant at the 5* level, (**)significant at the 1% level, (***)significant at the 0.1% level (d) This coefficient should be interpreted carefully since N is small and being treat was almost 1. Defendant can plead to misdemeanor or felony (violent or non-violent).

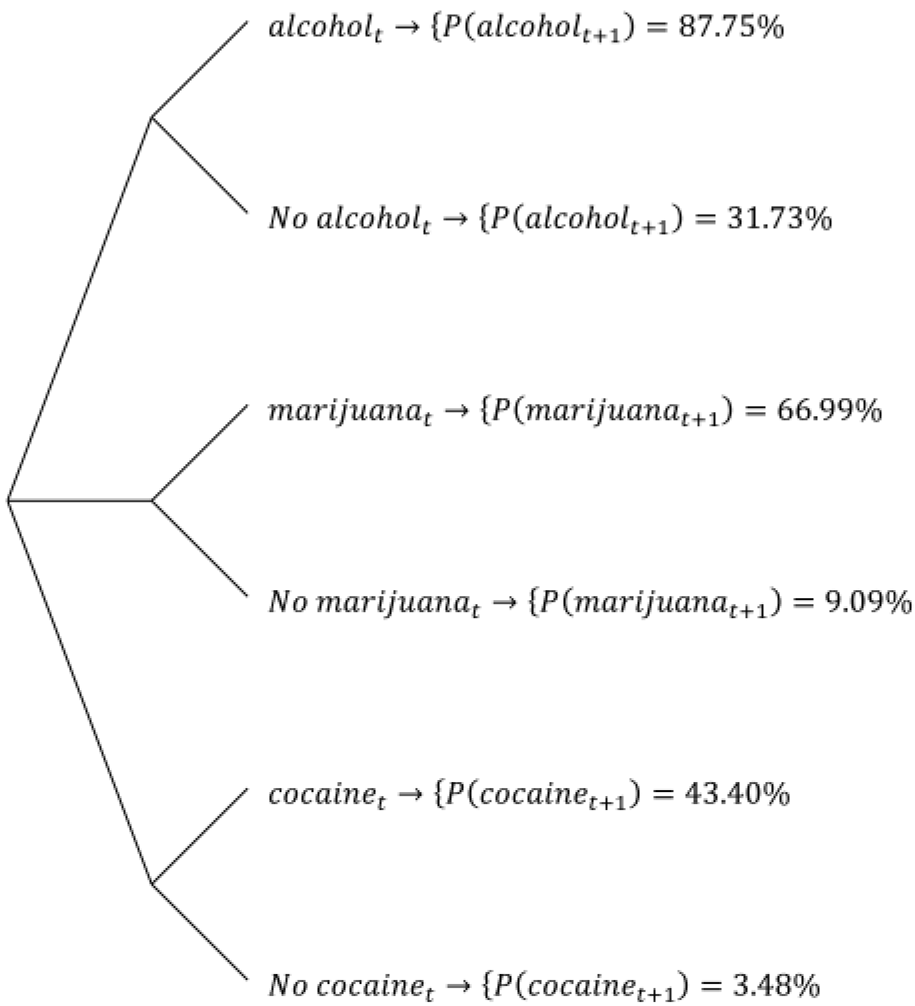

Figure 1.1: Drug Persistence

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 1.2: “Transitions” from Softer to Harder Drugs

Note: I computed these probabilities for every year starting in 1999, and then took the average.

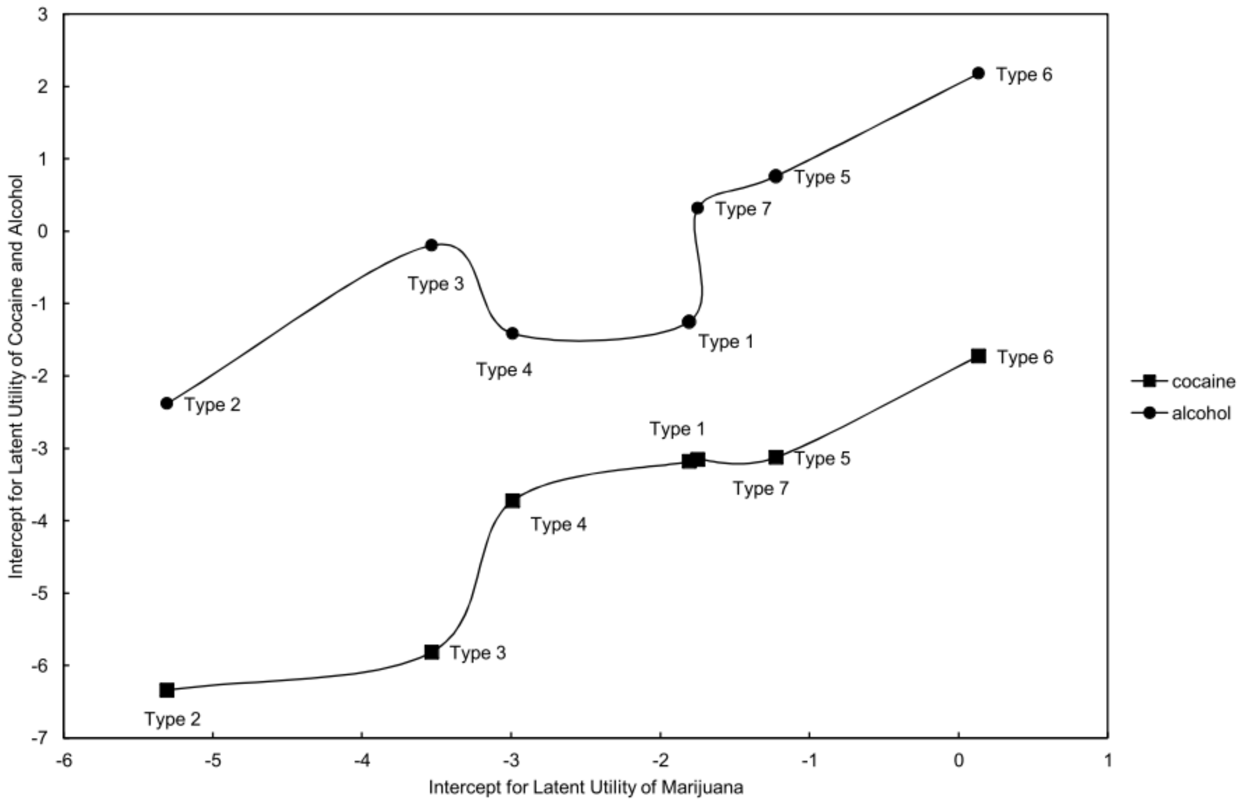

Figure 1.3A: Arbitrarily correlated unobserved heterogeneity (Model A)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.

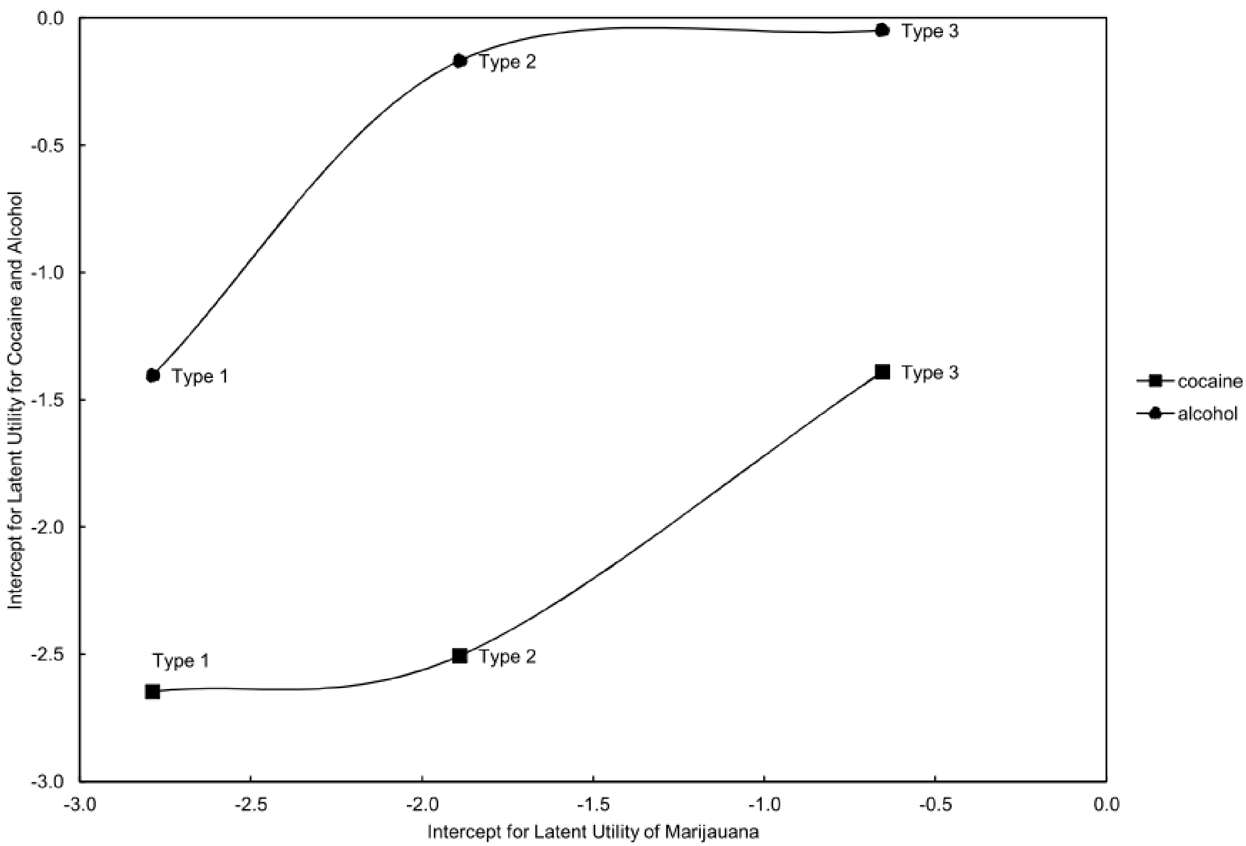

Figure 1.3B: Arbitrarily correlated unobserved heterogeneity (Model B)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.

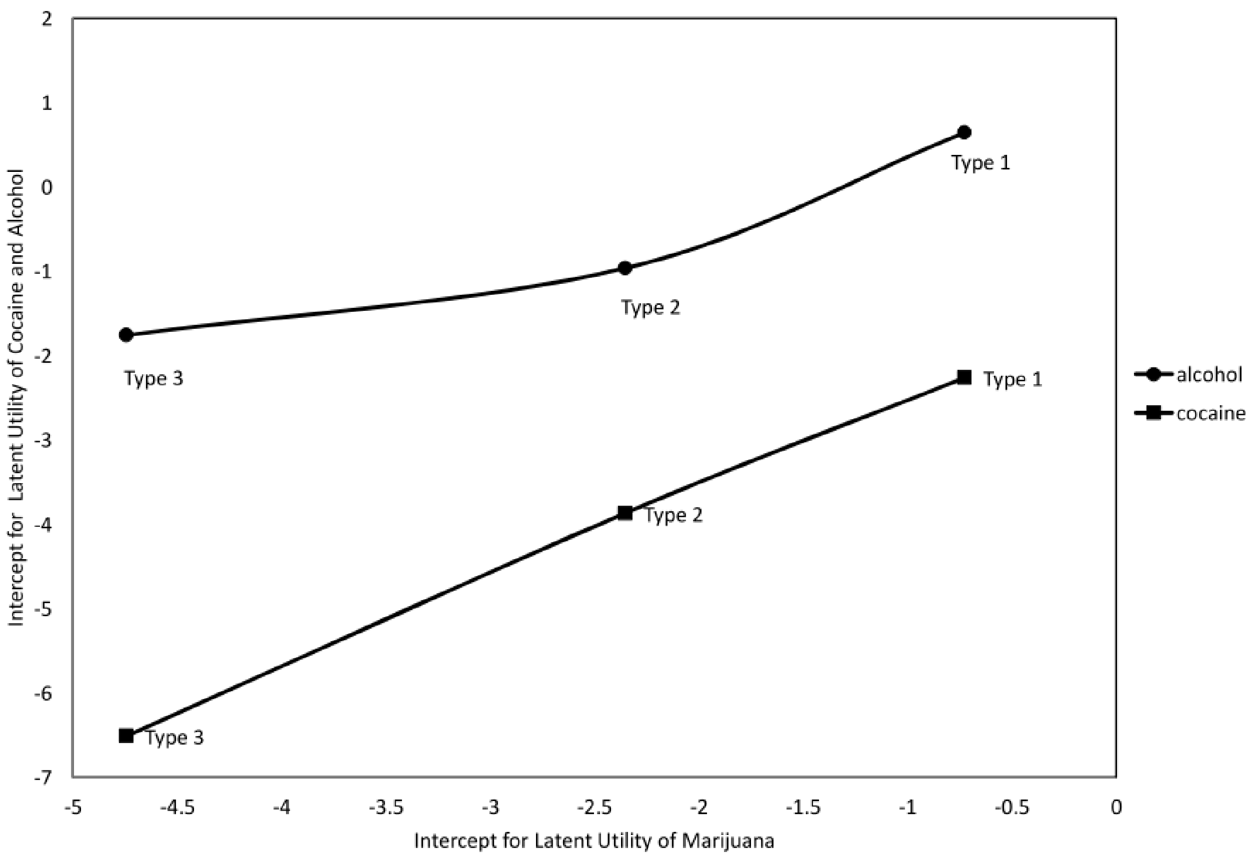

Figure 1.3C: Arbitrarily correlated unobserved heterogeneity (Model C)

Note: This figure plots the intercept corresponding to the marijuana latent utility against the one corresponding to the alcohol and cocaine latent utility.

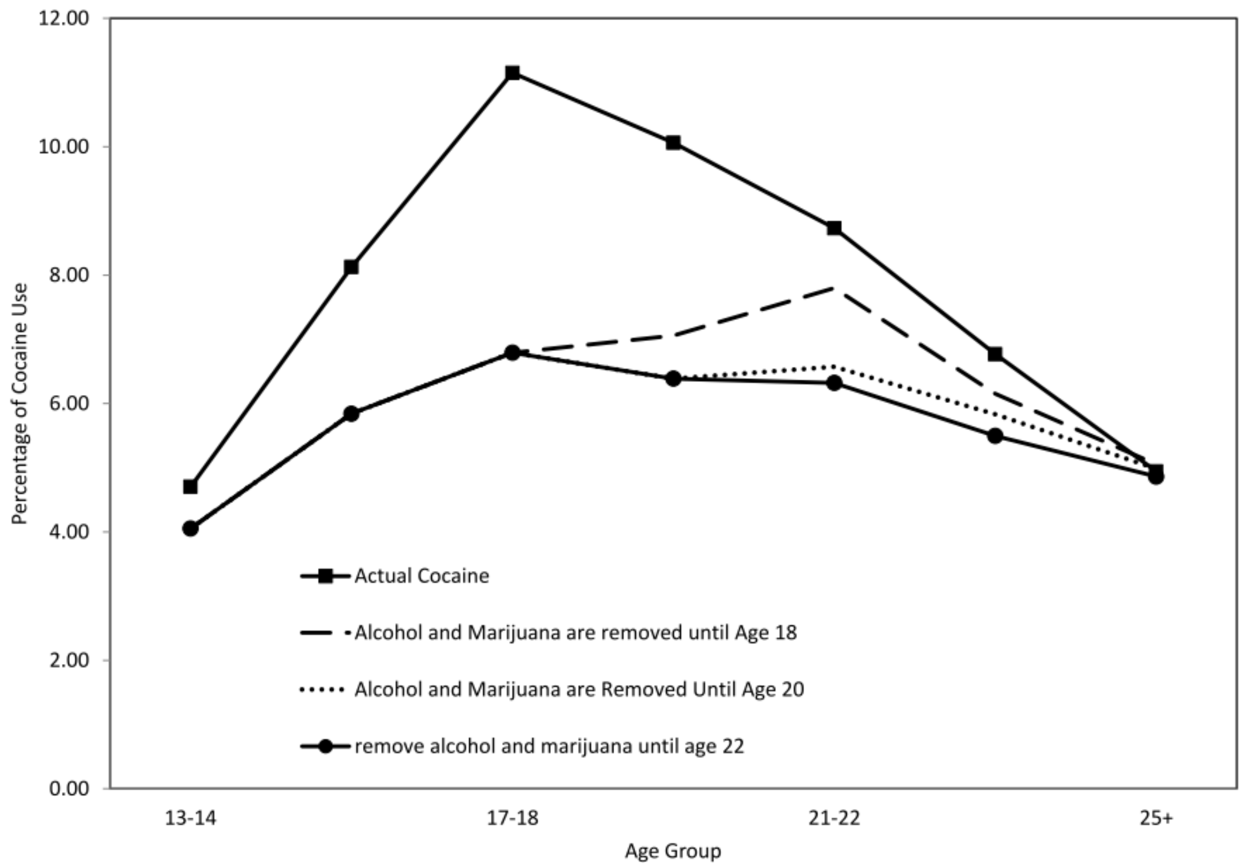

Figure 1.4: Simulated impacts of preventing drug use at early ages

Note: Age profile of cocaine consumption, and the simulated age profile under the assumption that alcohol and marijuana are completely removed at early ages.

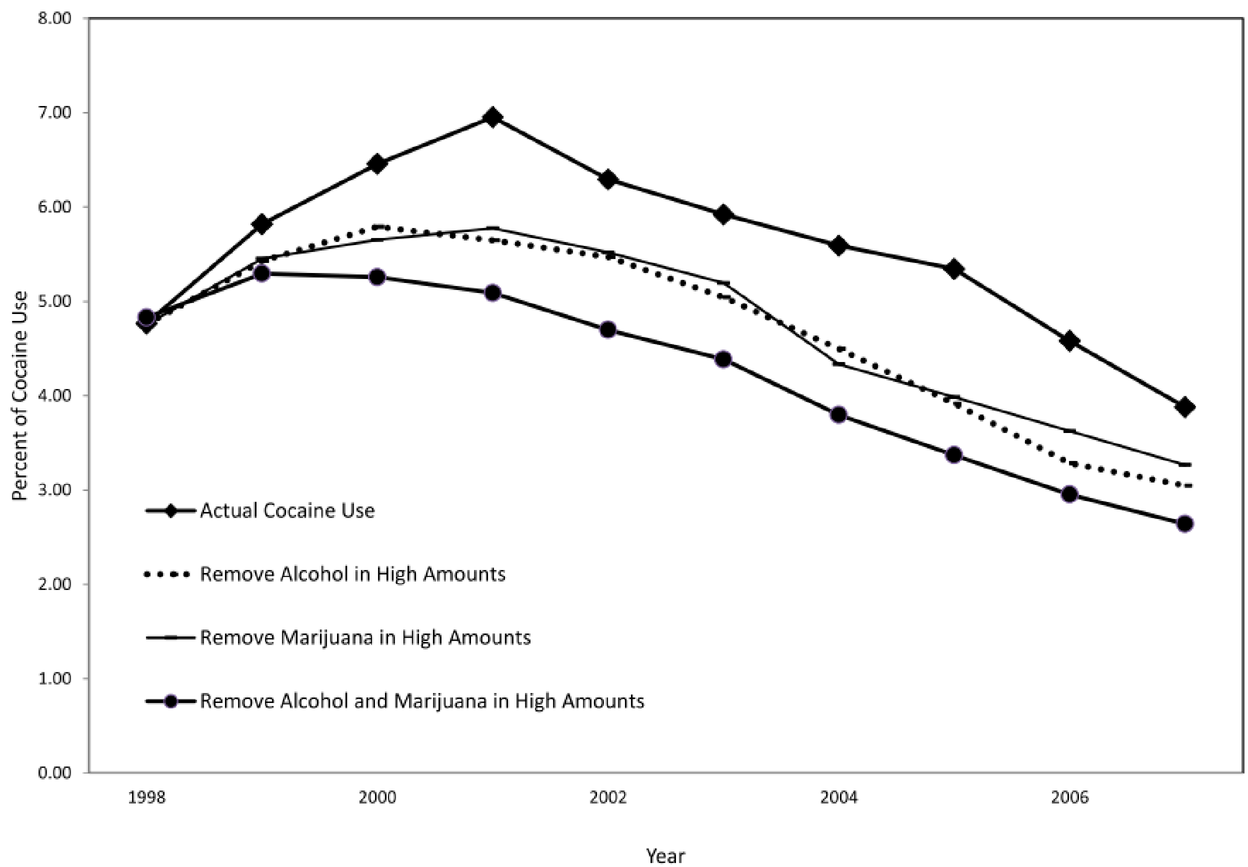

Figure 1.5: Simulated impacts of regulating intensity of drug use

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.

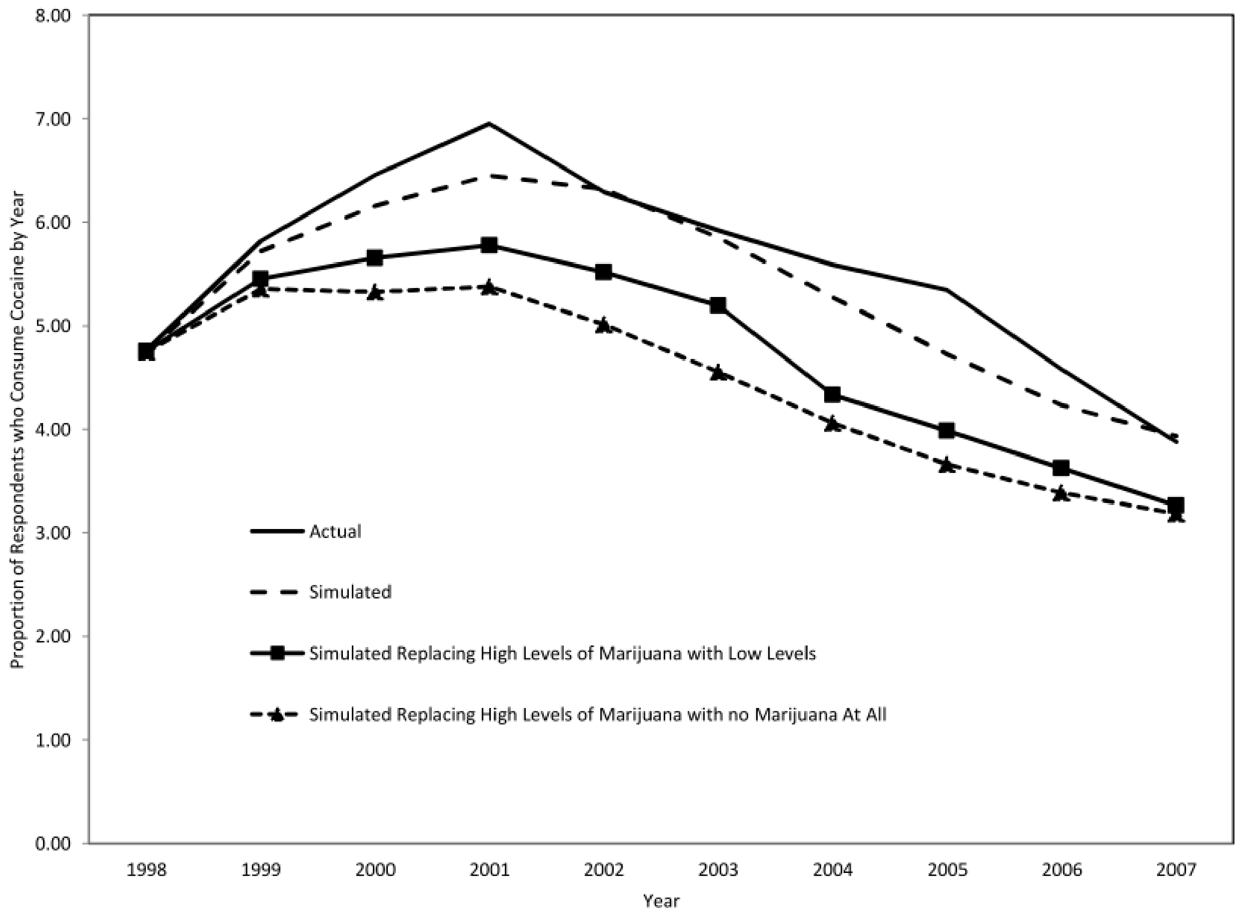

Figure 1.6: Simulated impacts of restrictions on marijuana intensity of use

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.

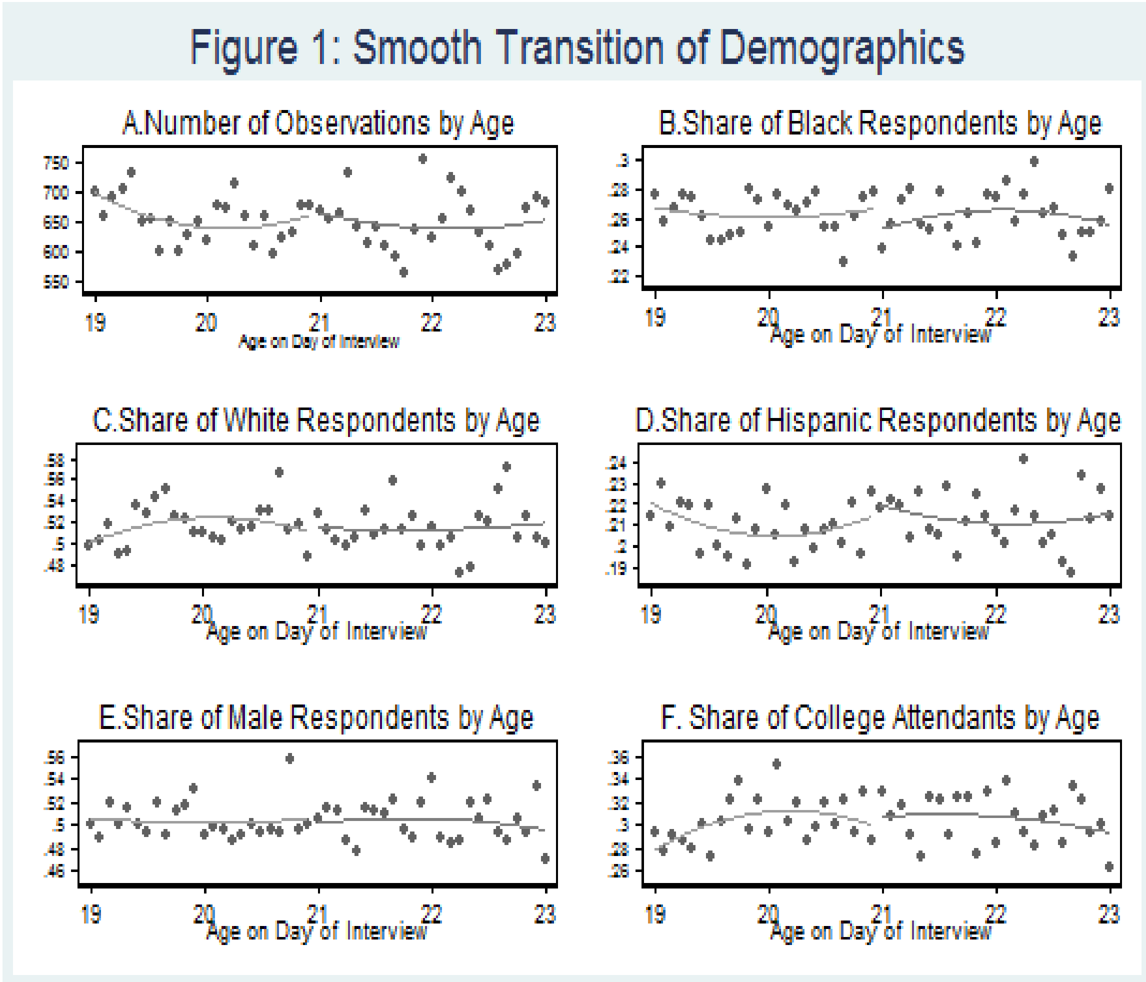

Figure 2.1 Smooth Transition of Demographics

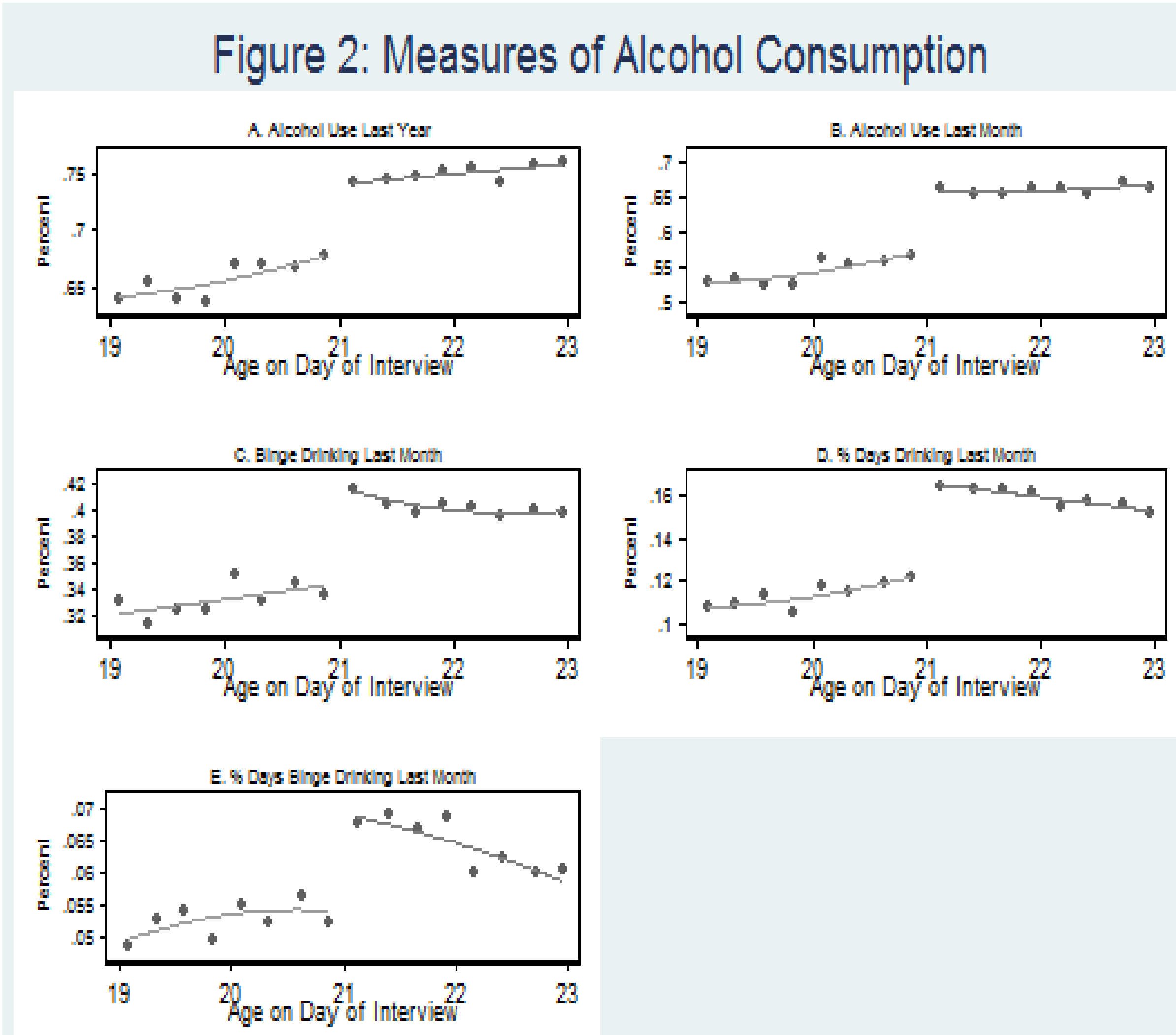

Figure 2.2 Measures of Alcohol Consumption

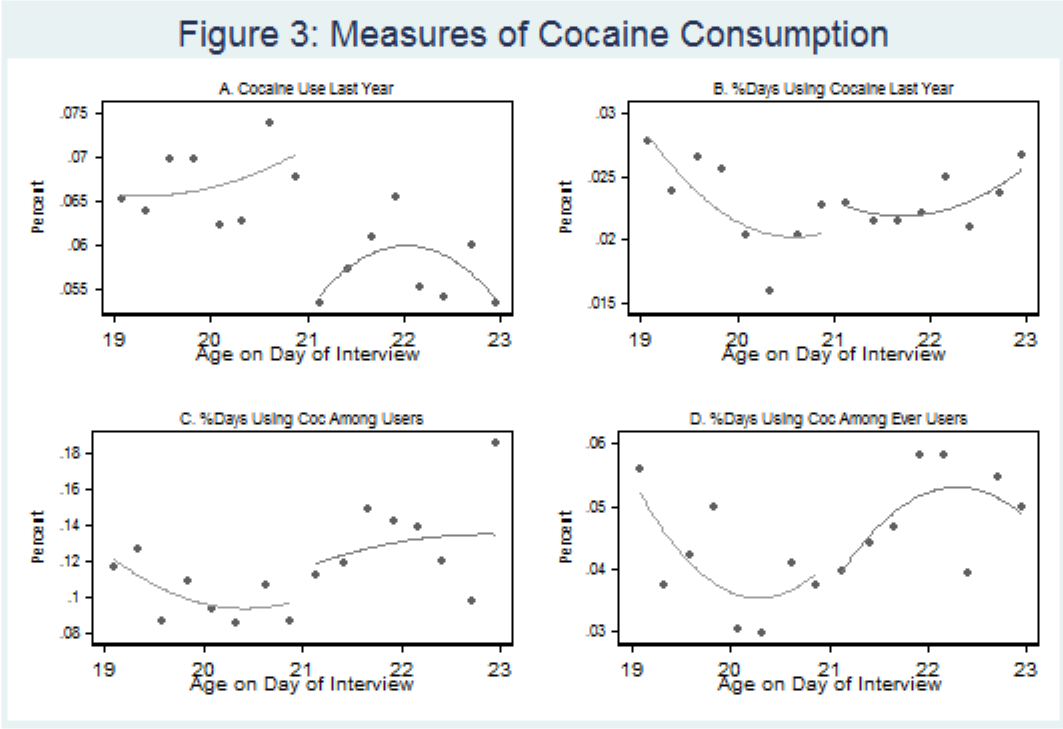

Figure 2.3 Measures of Cocaine Consumption

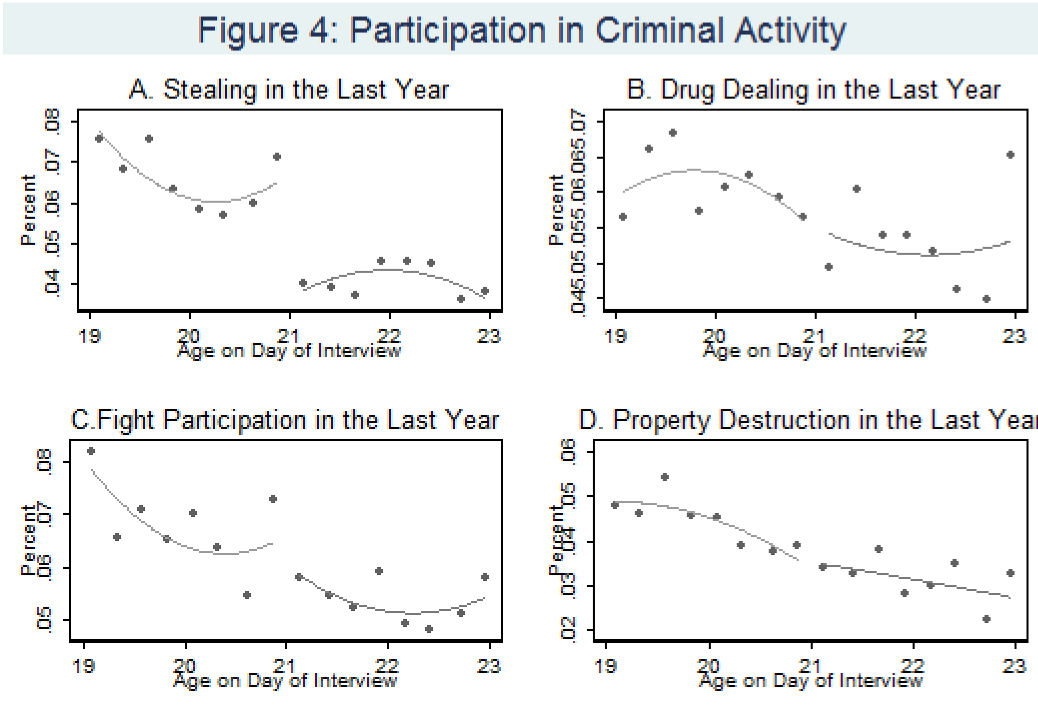

Figure 2.4 Participation in Criminal Activity

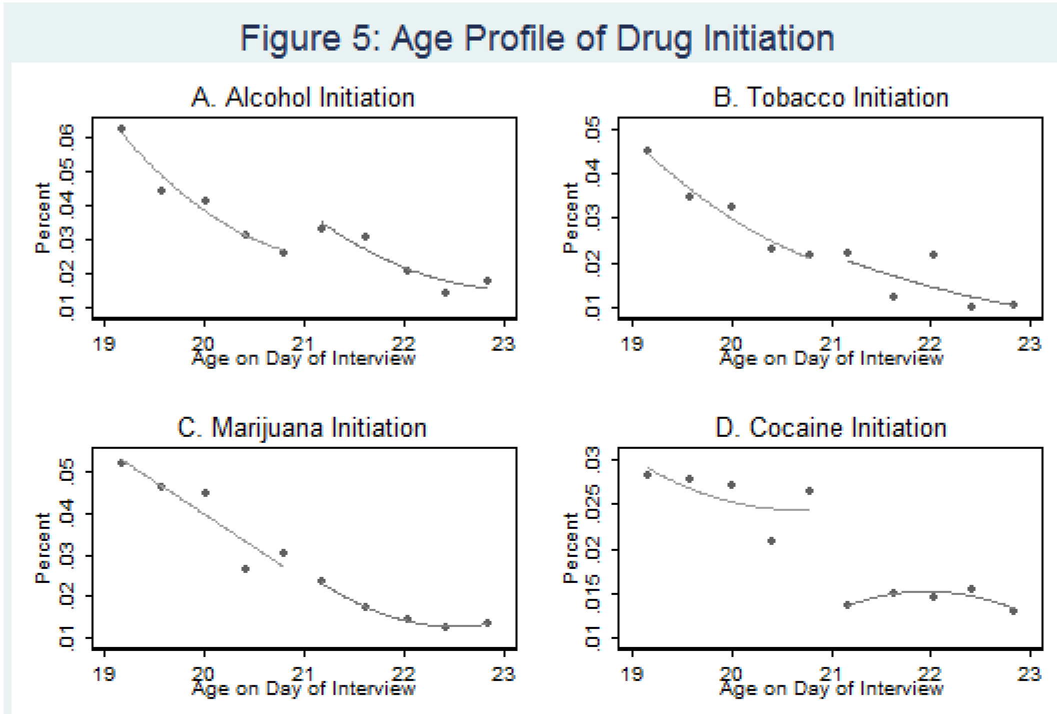

Figure 2.5 Age Profile of Drug Initiation

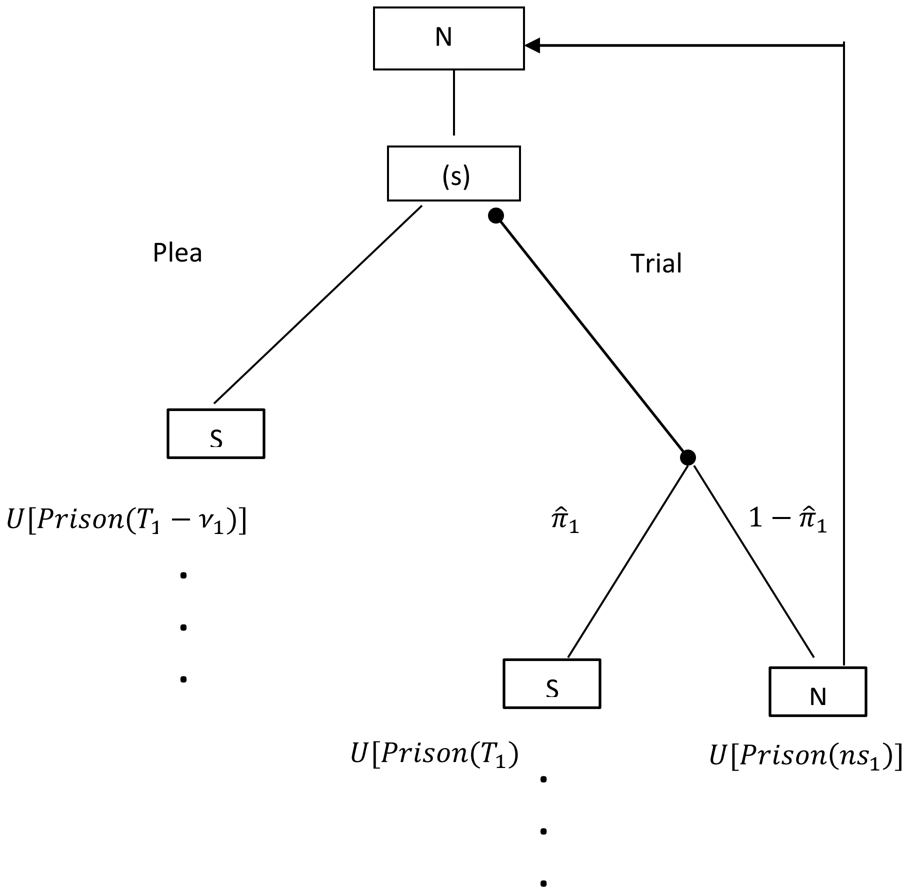

Figure 3.1. Decision Tree for Defendants Arrested for a Potential First Strike

Note: The three dots indicate that S is not a final state.

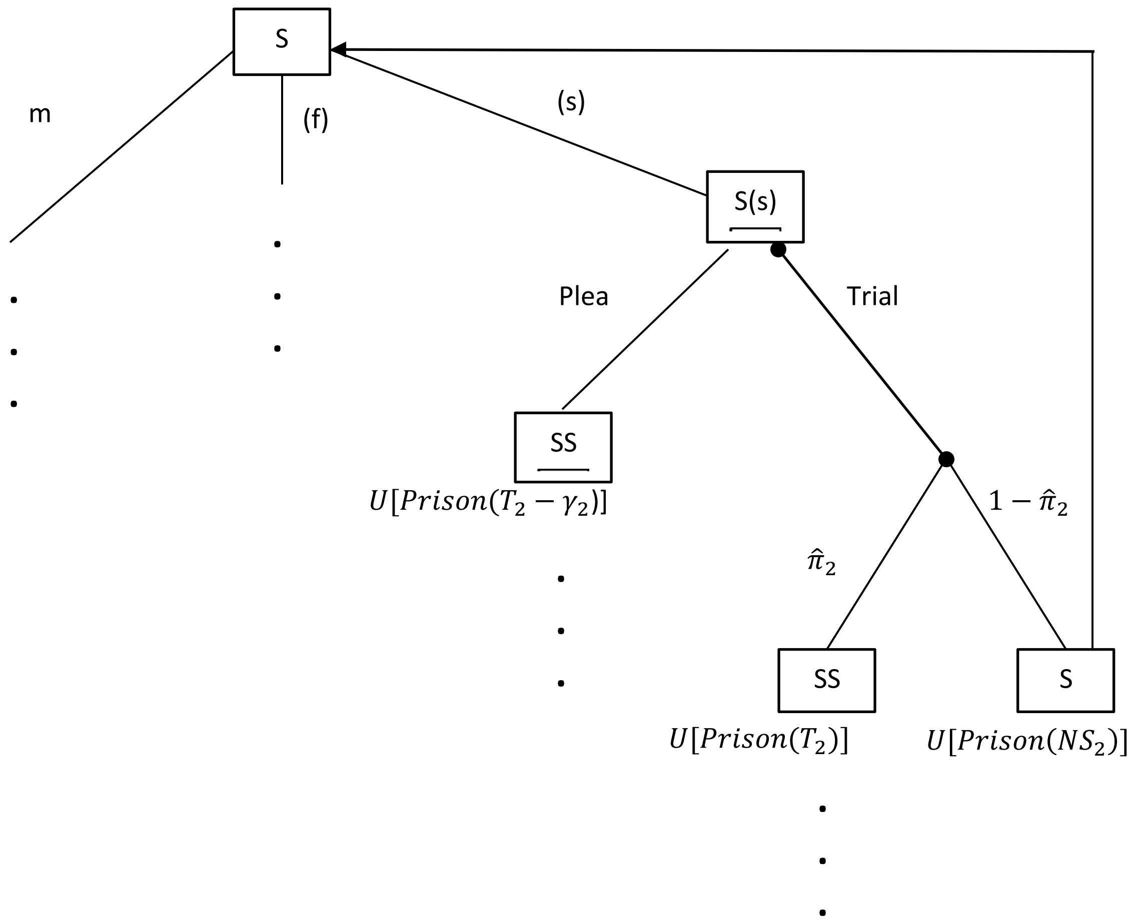

Figure 3.2 Decision Tree for Defendants Arrested for a Potential Second Strike

Note: The three dots indicate the continuation value of being at each state

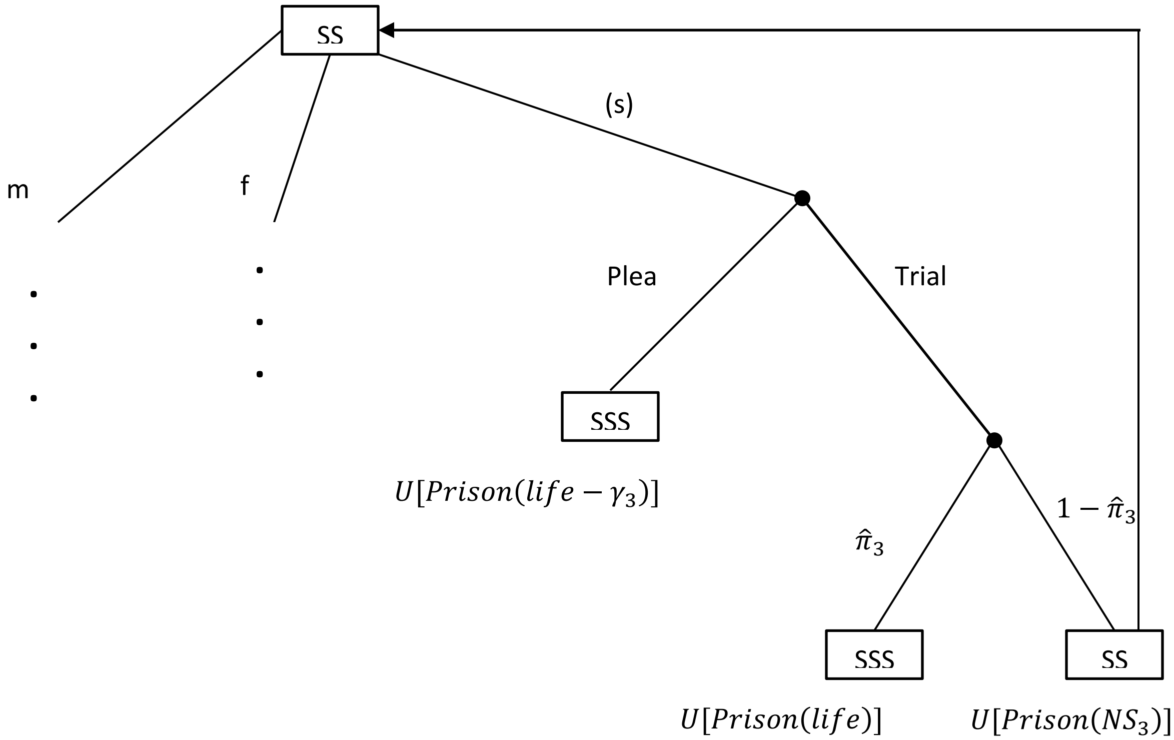

Figure 3.3. Decision Tree for Defendants Arrested for Potential Third Strike

Note: The three dots indicate that the state is not terminal .

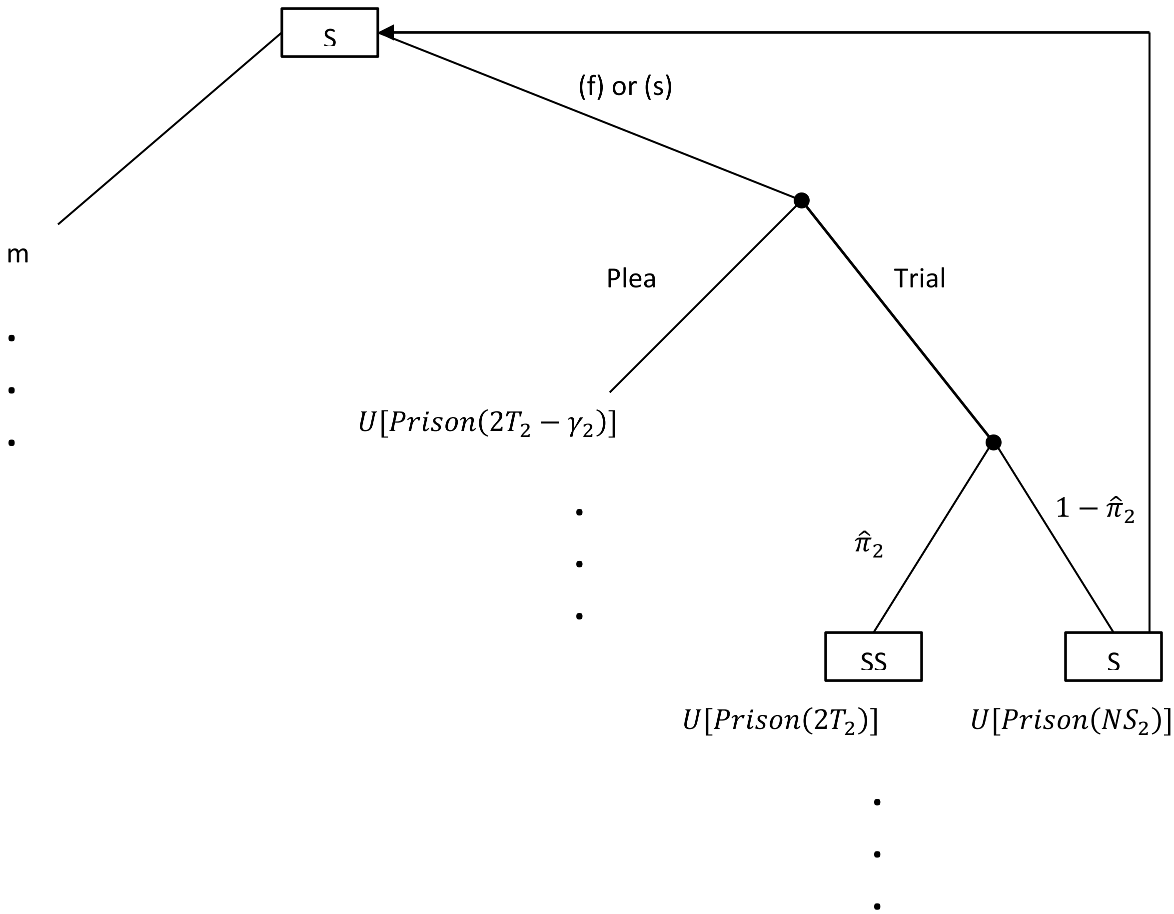

Figure 3.4 Decision Tree for Defendants Arrested for a Potential Second Strike in California

Note: The three dots indicate the continuation value of being at each state