High Today Versus Lows Tomorrow:

Substance Use, Education, and Employment Choices of Young Men

Catherine Alford*

October 28, 2014

Abstract

In this paper, I develop and estimate a dynamic structural model of employment, educational, and substance use decisions of young men in order to determine the causal effects of substance use on educational attainment and career paths. Heavy substance use is correlated with lower school attainment and labor market outcomes; however, it is unclear if heavy substance use *causes* these worse outcomes. Variation in the prices of substances, college prices, law enforcement characteristics, and unemployment rates help identify the channels through which current substance use and schooling decisions affect future substance use, employment decisions, and wages. Current research generally treats the substance use decision as a binary choice, making it difficult to distinguish the effects of moderate versus heavy use. I allow individuals to make choices about their level of alcohol, cigarettes, and marijuana use. I use restricted-access 1997 National Longitudinal Survey of Youth data and estimate my model with a Bayesian Markov Chain Monte Carlo algorithm. In preliminary results, I find that exogenous substance use at age 15 would substantially increase the probability of being a heavy substance user at age 24. If the substance use at age 15 is moderate, it has no effect on high school graduation rates at the age of 24. On the other hand, I find that exogenous heavy substance use at age 15 decreases the probability of high school graduation of white and Hispanic males by 5.62 and 7.59 percent, respectively.

JEL Classification: J24, J31, C11, I12

Keywords: Substance Use, Human Capital Accumulation, Education, Labor Productivity, Wage, Bayesian Estimation

---

*Ph.D. candidate in the Economics Department at the University of Virginia. Email: cea9e@virginia.edu. I would like to thank my advisors, Steven Stern, Leora Friedberg, and Sarah Turner for their invaluable time and advice. I would also like to thank Andrew Ching, Victor Aguirregabiria, and Aloysius Siow and participants at the SWEAT workshop at the University of Toronto for their helpful suggestions. I would also like to thank Matthew Harris for his comments. I am grate for the countless conversations with several of my classmates. Lastly, I would like to thank the University of Virgina's Bankard Fund for financial support. All remaining errors are my own.

1 Introduction

Substance use is prevalent among American youth. In 2012, 6.8 percent of 12 to 17 year old males reported smoking cigarettes in the past month, 12.6 percent reported drinking, and 7.5 percent reported using marijuana.[1] Heavy substance use is correlated with lower school attainment and labor market outcomes; however, it is unclear if heavy substance use *causes* these worse outcomes. One concern is that those who are more likely to use, say, marijuana frequently are those for whom the labor market would offer lower wages regardless of marijuana use. It may also be the case that poor expected labor market outcomes make substance use less costly.

In this paper, I develop a model that captures these channels relating substance use, educational attainment, and career paths of young men. I develop a dynamic structural model where individuals make decisions about schooling and work as well as how much alcohol, cigarettes, and marijuana to consume. My model allows substance use to affect career paths through its effects on educational attainment, criminal record, and wage offers. I allow substance use to affect whether an individual enrolls in school as well as whether an individual advances to the next grade if he enrolls. Additionally, substance use in my model affects whether an individual gets arrested, whether an individual chooses to work, and wages. Individuals make decisions in order to maximize their discounted lifetime utility, accounting for the effects of substance use on future outcomes. Individuals are observed making different choices due to both observed characteristics, such as prices and legality of substance use, and unobserved characteristics, such as their individual predilection for substance use.

A large literature, both reduced-form and structural, examines how substance use affects educational attainment and wages. Typically, research on this subject focuses on only one substance and one outcome. However, youth who engage in one form of risky behavior are likely to participate in other forms as well. For example, consider the finding that youth who smoke cigarettes have lower educational attainment than those who do not. If individuals who smoke cigarettes do so only when they consume alcohol, then perhaps it is the alcohol that causes the lower educational attainment and not the cigarettes. Moreover, the current literature generally limits the substance use decision to a binary choice. Yet, it is commonly believed that heavy substance use adversely affects education and career outcomes more than moderate use does. I allow individuals to make choices about their level of substance use to capture the full relationship between substance use levels and outcomes.

I estimate my model using Bayesian Markov Chain Monte Carlo (MCMC) methods. My model includes extensive unobserved heterogeneity and continuous state variables, which makes evaluating the likelihood function computationally difficult. I use a modified version of the estimator

[1] 2012 National Survey on Drug Use and Health.

proposed in Imai, Jain, and Ching (2009) (IJC) in order to make estimation feasible by easing the computational burden of evaluating my likelihood and value functions. I use restricted-access data from the 1997 National Longitudinal Survey of Youth (NLSY97) to estimate my model. The NLSY97 collects information about labor market behavior and educational experiences as well as information on drug use. Variation in the prices of substances, college prices, law enforcement characteristics, and unemployment rates help identify the channels through which current substance use and schooling decisions affect future substance use, employment decisions, and wages. Most papers in the literature use marijuana prices from the Drug Enforcement Administration's (DEA's) System to Retrieve Information from Drug Evidence. The DEA's focus is on harder drugs like heroin, so the number of marijuana observations are small, and a large amount of those observations are concentrated in Washington D.C.. I use marijuana price data assembled from High Times magazine, which has much better geographic coverage. Details of how I assemble these prices can be found in Alford (2013).

In preliminary results, I find that the effects of substance use on outcomes vary across race, the type of substance used, and the amount used.[2] I find that consuming heavy amounts of marijuana decreases the utility of enrolling in school for all males; consuming heavy amounts of alcohol decreases the utility of enrolling for black males but increases it for Hispanic males; and consuming heavy amounts of cigarettes decreases the utility of enrolling for Hispanic and black males. Additionally, heavy substance use decreases the probability of passing a grade for all males. For all races, heavy substance use increases the utility of working, perhaps through its effect on the budget constraint. Heavy cigarette use decreases wages of Hispanic males by 26.63 percent, while heavy marijuana use increases wages for white males. Additionally, I find that cigarette, alcohol, and marijuana use are complements. This suggests that policies aimed at reducing the use of one substance will decrease the use of all three substances. Lastly, I find that past alcohol use has no direct effect on current marijuana use for white males, but increases the utility of using marijuana for black and Hispanic males. This implies that alcohol is a gateway drug for minority males.

I find that unobserved heterogeneity in preferences and skills also vary across race. For example, white males with a higher than average preference for heavy marijuana use also have higher preferences for working, higher chances of getting arrested, higher chances of passing a grade, and lower earnings. Hispanic males with higher preferences for heavy marijuana use have higher preferences for working part-time and lower preferences for enrolling in school. Black males with higher preferences for heavy marijuana use have higher preferences for working part-time, lower preferences for enrolling in school, lower chances of passing if enrolled in school, and lower chances of getting arrested.

[2] To get these results, I simplify some relatively minor parts of the stochastic and functional form specifications and estimate the model on 150 males from each race (450 males total).

In order to interpret the results presented above, I run a counter-factual simulation to investigate the effects of exogenously making someone a moderate or heavy substance user at age 15. I find that exogenous substance use at age 15 substantially increases the probability of being a heavy substance user at age 24. Moderate substance use at age 15 has no effect on high school graduation rates or wages at the age of 24. On the other hand, I find that exogenously using a heavy amount of substances at age 15 decreases the probability of high school graduation of white and Hispanic males by 5.62 and 7.59 percent, respectively. I only observe most individuals in my sample until the age of 24, which may be causing me to underestimate the returns to education in the wage equation. Therefore, although I estimate lower high school graduation rates for white and Hispanic males who exogenously use heavy amounts of substances, I find there is no statistically significant effect of heavy use on log-wage offers at the age of 24. I am currently running counter-factual simulations to show, for example, the effects of changes in substance prices or the probability of arrest on substance use, education, and employment decisions.

2 Literature Review

My paper builds on previous work in two areas of the literature. First, my paper is motivated by the substance use literature, which includes papers that analyze the determinants of substance use, and addiction and papers that analyze the effects of use on labor market and educational outcomes. Second, my model expands on those developed in the human capital investment literature, particularly involving dynamic labor supply, by modeling leisure decisions of individuals.

2.1 Substance Use Literature

Rational Addiction Literature: Most empirical studies of the rational addiction theory follow Becker and Murphy (1988). They develop a theory in which individuals make decisions about consuming goods that they know to be addictive by weighing how their choices today will affect their future utility. Chaloupka (1991) and Becker, Grossman, and Murphy (1994) were two of the first papers to empirically test this theory and find evidence supporting it. Labeaga (1999) improves on this by estimating a rational addiction model for tobacco that incorporates serial correlation and unobserved heterogeneity. The empirical results reject the myopic model and give support to the rational addiction model. Gruber and Köszegi (2001) also find evidence supporting forward-looking behavior but argue that individual preferences may be time-inconsistent. Arcidiacono, Sieg, and Sloan (2007) develop a different test of the rational addiction model. Rather than seeing if substance use responds to prices, they explore whether substance use of the near-elderly responds to negative health and income shocks. They similarly find that forward-looking models fit the data better than myopic models. This literature motivates my decision to model individuals as rational, forward-looking decision makers.

Substance Use Effects on Educational Outcomes: Several authors have studied the effects of substance use on educational outcomes such as graduation and college enrollment. There is no clear consensus on how substance use affects educational attainment. Cook and Moore (1993) find that students who attend high school in states with high beer taxes or a high minimum drinking age complete more years of schooling. Yamada, Kendix, and Yamada (1996) find significant adverse effects of current alcohol and marijuana use on high school graduation rates. Alternatively, Dee and Evans (2003) conclude that alcohol consumption of teenagers does not reduce educational attainment. Bray et al. (2000) and Chatteriji (2006) find evidence that marijuana use decreases the probability of high school graduation.

In my model, I am not relying solely on variation in state laws or prices to determine an individual's substance use. Individuals are also influenced by shocks that affect their educational attainment, wages, and job opportunities. I also model the usage of several substances, allowing me to control for the interdependencies of cigarette, alcohol, and marijuana use. By explicitly modeling substance use decisions within a human capital accumulation model, I can determine the channel through which substance use affects educational attainment. That is, I can determine whether lower achievement among users is due to decreased enrollment, increased probability of failing, or because individuals who use substances simply do not enjoy school. Lastly, the estimates from my model allow me to conduct counter-factual simulations, for example, of the effect of changes in substance prices and legality on educational attainment and employment decisions.

Substance Use Effects on Labor Market Outcomes: Several studies have examined the effects of substance use on labor market outcomes, such as wages, employment, and hours of work. Often, authors use past substance use, rather than current substance use, in their regressions, which they argue is independent of the error term (Buchmueller and Zuvekas 1998, Zarkin et al. 1998). DeSimone (2002), on the other hand, instruments for current substance use and finds that the use of marijuana and cocaine each substantially reduces the likelihood of employment. Kaestner (1994) also treats substance use as endogenous and, after controlling for unobserved heterogeneity, he finds no effect of illicit drug use on labor supply. All of the papers listed above treat education as an exogenous determinant of labor market outcomes.

Bray (2005) improves on this earlier work by treating alcohol use, educational attainment, and work experience as endogenous. In his model, he allows educational attainment, work experience, and alcohol use to affect wages. He models the enrollment, work, and alcohol use decisions as logit models that are designed to capture the correlation of the decisions with the error term in the wage equation. The reduced-form equations are auxiliary estimating equations and, therefore, cannot be directly interpreted. Estimation results suggest that moderate alcohol use while in school or working has a positive effect on human capital accumulation, and heavier drinking reduces this gain. My model differs from Bray's in two ways. First, I develop and estimate a structural model of individuals' schooling, work, and substance use decisions, which allows me to see how current substance use affects not only wages but also schooling decisions and future substance use. Second, individuals in my model make decisions about alcohol, cigarette, and marijuana use, which allows me to account for the interdependencies of substance use.

2.2 Human Capital Investment and Dynamic Labor Supply Literature

Most papers on human capital investment are based on Roy (1951) and Heckman and Honore (1990), which investigate how selection into occupations affects the distribution of earnings and productivity in those occupations. Eckstein and Wolpin (1999) use this framework to look at youths' high school dropout decisions. Their findings indicate that dropping out of high school is confined to youths with lower ability, a lower expected value of a high school diploma, a comparative advantage in skills suited for jobs that do not require a high school diploma, and a lower consumption value of attending school. Whereas Eckstein and Wolpin (1999) mainly focuses on educational outcomes, another set of papers focuses more on labor market outcomes. Keane and Wolpin (1997) and Sullivan (2010) examine dynamic educational and occupational choices and allow work experience and education to be accumulated endogenously. In Keane and Wolpin (1997), individuals choose whether to attend school, to work in one of three occupations, or to engage in home production. Sullivan (2010) expands this choice set by allowing individuals to select into more than three occupations, to select into firms in addition to occupations, and to participate in dual activities such as employment while attending school. Following Sullivan (2010), I allow individuals to select into firms, and I allow for dual activities. Unlike Sullivan (2010) and Keane and Wolpin (1997), I allow individuals to pick part-time or full-time hours, which is particularly important for individuals while they are enrolled in school.

Mezza (2011) is most closely related to my paper. He estimates a dynamic structural model where individuals jointly make decisions about whether to consume drugs, attend school, and participate in the labor force. He finds that non-drug users have higher wages than marijuana and/or hard drug users. My model differs from the model presented in Mezza (2011) in three main ways. Most importantly, I allow individuals to choose different levels of substance use. Secondly, I allow substance use to affect the probability that an individual gets arrested and allow arrests to affect human capital accumulation choices. This allows me to determine whether it is the use itself that is affecting outcomes or if the effect is coming through arrests. Lastly, my focus is on softer drugs, that is cigarette, alcohol, and marijuana use, whereas his focus is on harder substance use. Cigarette and alcohol use may have strong gateway effects among youth that Mezza (2011) is not able to capture. These gateway effects could have strong policy implications for combating youth substance abuse.

3 Model

In this section, I show how young men make decisions about schooling, employment, and substance use. Some modeling choices, especially about substance use, are heavily influenced by data availability, which is discussed in more detail in Section 4. I use a finite horizon, discrete time, dynamic model. In each year, from age 14 ($t = 1$) to a known terminal age ($t = T$), an individual i makes decisions in order to maximize the discounted sum of his lifetime utility subject to annual budget constraints. An individual makes choices about his substance use as well as about his employment and education. He chooses whether to consume no cigarettes, a moderate amount of cigarettes, or a heavy amount of cigarettes, $cigs_{it} \in [1, 2, 3]$; how many days to consume alcohol (none, moderate, or heavy), $alc_{it} \in [1, 2, 3]$; and how many days to consume marijuana (none,moderate, or heavy), $mj_{it} \in [1, 2, 3]$. Let $sub_{it} = (cigs_{it},\ alc_{it},\ mj_{it})$ denote the substance use choices. The human capital accumulation choices individuals make are: whether to enroll in school, $enroll_{it} = 1$; whether to be unemployed $h_{it} = 0$, work part-time $h_{it} = 1$, or work full-time $h_{it} = 2$; and, if working, whether to work for a new employer, $ne_{it} = 1$. I denote the vector of human capital accumulation choices as $hc_{it} = (enroll_{it}, h_{it}, ne_{it})$. An individual receives one new part-time offer and one new full-time offer each period. If he worked in the previous period, then he also has the choice to continue working at the same job. An individual can always choose not to work or to enroll in school. Let $K_{it}(s_{it})$ denote the set of human capital accumulation choices an individual faces, where s_{it} denotes the state that individual i is in at time t. The choice set that individual i faces, $K_{it}(s_{it})$, varies with s_{it} because individuals who worked last period have an additional discrete choice, whether to continue working at the same job .

Individuals are observed choosing different paths due to observed and unobserved heterogeneity that affect the utility associated with each choice. For example, someone may choose to work at a very low wage while enrolled in school because the experience and education will raise their wages in the future, while someone who has a lower preference for work might choose to not work. My model allows substance use to affect career paths through its effect on educational attainment, wage offers, and arrests. The effect of use on educational attainment is ambiguous, operating through enrollment and grade completion. School may make obtaining drugs easier, especially if other youths in the school are using drugs, which would encourage users to enroll in school. However, substance use could make it harder for a youth to succeed in school, as using substances may crowd out time spent on schoolwork and decrease cognitive ability. These effects raise the cost for a person using substances to put forth the same amount of effort towards school as a person not using and may discourage enrollment. Individuals in my model can only choose whether or not to enroll, not their educational attainment. Individuals who choose to enroll have a probability of passing that is a function of their substance use. Current substance use can also affect job opportunities by altering productivity and hence the wage offers the individual receives. Lastly, I allow substance use to affect whether or not an individual gets arrested. I allow arrests to affect the enrollment and work decisions as well as the wage offers.

3.1 Choice-Specific Utility Flows

In each period t, utility is comprised of pecuniary utility from the consumption of goods $CONS_{it}$ and nonpecuniary utility N_{it}. The utility an individual gets from making choice (hc_{it}, sub_{it}) is a function of endogenous state variables s_{it}, skill and preference endowments, and random shocks that vary over time, people, and employers. The endogenous state variables measure educational attainment, work experience, firm-specific human capital, arrest histories, and the addictive stocks of the three substances. The utility flow that individual i receives at time t from choices hc_{it} and sub_{it} is

$$u_{it}\left(hc_{it}, sub_{it} | s_{it}\right) = \alpha \ln\left(CONS_{it}\right) + (1-\alpha) \ln\left(N_{it}\right) \tag{1}$$

The remainder of this subsection describes the structure of the pecuniary and nonpecuniary utility flows.

3.1.1 Pecuniary Utility Flows

Assuming no savings, an individual's budget constraint equates his expenditures $CONS_{it}$ to his income.[3] Potential earnings from working h_{it} hours at firm k is denoted w_{itk}^{h}. It is not straightforward to measure and model income, consumption, and wealth of youths receiving support from their parents. Yet, youths make substance use and work decisions that take income into account. Therefore, I allow individuals to receive transfers $\widetilde{w}_{it}\left(hc_{it}\right)$ from an outside source, for example, family members or the government.[4] Expenditures include purchases of a composite commodity with price equal to one; purchases of cigarettes at price p_{it}^{c}, alcohol at price p_{it}^{a}, and marijuana at

[3] I rule out saving for college or to purchase substances as a motivation for working while in high school. Modeling savings would require using accumulated earnings as a state variable that affects school and substance use decisions. However, it is not obvious how to measure wealth for a youth still living with their parents.

[4] Similarly to Eckstein and Wolpin (1999), I do not model the parents' decision about how much to transfer to their children, so $\widetilde{w}_{it}$ is treated as exogenous. I allow the distribution of transfers to vary by discrete choice to capture the fact that an individual not working and enrolled in college needs more (and may be rewarded with more) transfers than an individual not going to school and working full-time. This is to avoid having to model game-theoretic interactions between youths and their parents, as in Rosenzweig and Wolpin (1994); Hao, Hotz, and Jin (2008); and Martinez (2014)

price p_{it}^{m}; and expenses p_{it}^{s} associated with attending college. Thus, in each period, an individual's budget constraint is

$$CONS_{it} + p_{it}^{sub} sub_{it} + p_{it}^{s} \mathbb{1}\left(G_{it} > 12\right) \mathbb{1}\left(enroll_{it} = 1\right) = w_{it}^{h} h_{it} + \widetilde{w}_{it} \tag{2}$$

where $p_{it}^{sub} = \left(p_{it,}^{c}\, p_{it}^{a},\, p_{it}^{m}\right)$ and G_{it} is the educational attainment at time period t. I assume transfers are large enough so that consumption is above a minimum level, C_{min}.

Individuals who work at the same firm for several years are able to build firm-specific human capital that raises their productivity. Individuals in my model are allowed to choose to work at the firm they worked at in the previous period or at a new firm. An individual's log-wage at firm k in period t is

$$\ln\left(w_{itk}^{h}\right) = \sum_{h=1}^{2} \theta_{w}^{h}\left(sub_{it}, s_{it}\right) \mathbb{1}\left(h_{it} = h\right) + \psi_{ik} + \mu_{i}^{w} + \nu_{itk}^{h}. \tag{3}$$

The term $\theta_{w}^{h}\left(sub_{it}, s_{it}\right)$ represents the deterministic part of the log-wage, which is a function of human capital, arrest history, and today's substance use. This specification allows the effects of state variables and substance use on the wage to differ by hours choice. ψ_{ik} is a permanent worker-firm productivity match value. This reflects factors, such as an individual's rapport with his boss, which affect the wage of worker i at firm k. Time-persistent individual heterogeneity in productivity that does not vary over time nor jobs is denoted μ_{i}^{w}. The last term ν_{itk}^{h} is an idiosyncratic shock that captures true randomness in wages. Individuals observe all of the components of the wage when a job offer is received. Future realizations of firm-specific match values and wage shocks are unknown to the individual until they occur.

3.1.2 Nonpecuniary Utility Flows

Nonpecuniary utility is a function of individual characteristics and state variables. Numerous forms of state dependence may affect the ease or difficulty with which one can change states; therefore, nonpecuniary utility may differ according to one's choices in the previous period. For example, it is more difficult for a person to go to school after being un-enrolled. Also, tenure in a state may affect nonpecuniary utility. For example, it may be costly for an individual who has worked at the same firm for several periods to move to a new firm. Allowing tenure in substance use states to enter into the utility function also allows me to measure the addictiveness of substances. The log-nonpecuniary utility flow in period t is

$$
\begin{aligned}
\ln\left(N_{it}\right) &= n\left(hc_{it}, sub_{it}, s_{it}\right) \\
&+ \mu_i^{sub} sub_{it} + \mu_i^{h} \mathbb{1}\left(h_{it} = h\right) + \mu_i^{s} \mathbb{1}\left(enroll_{it} = 1\right) \\
&+ \epsilon_{it}^{hc} + \epsilon_{it}^{sub}
\end{aligned} \tag{4}
$$

where $\mu_i^{sub} = \left(\mu_i^{cigs=2}, \mu_i^{cigs=3}, \mu_i^{alc=2}, \mu_i^{alc=3}, \mu_i^{mj=2}, \mu_i^{mj=3}\right)$. The first line in equation 4 represents the deterministic part of the nonpecuniary utility that is a function of the choices and the state vector. This term includes the nonpecuniary utility flows an individual receives from employment, unemployment, and substance use as well as the cost function of attending school. The deterministic part of nonpecuniary utility is

$$
\begin{aligned}
n\left(hc_{it}, sub_{it}, s_{it}\right) &= \sum_{h=0}^{2} \theta_N^h \left(sub_{it}, s_{it}\right) \mathbb{1}\left(h_{it} = h\right) \\
&+ \kappa_S \left(h_{it}, sub_{it}, s_{it}\right) \mathbb{1}\left(enroll_{it} = 1\right) \\
&+ \alpha\left(sub_{it}, s_{it}\right).
\end{aligned} \tag{5}
$$

The first line of equation 5 captures the nonpecuniary utility of employment. This specification allows the effect of state variables and substance use on employment utility to vary by hours choice h. For example, an individual may get more enjoyment out of working full-time relative to not working if he is not consuming marijuana. This term also captures the utility gained from the leisure time consumed when unemployed. The second line is the nonpecuniary cost of enrolling in school. The cost function is allowed to differ depending on employment and substance use because it may be more difficult to attend school while employed full-time rather than part-time or when drinking heavily. The third line captures the utility gained from using substances.

The second line in the non-pecuniary utility in equation 4 captures time-persistent unobserved heterogeneity in preferences for substance use, hours worked, and school enrollment. Including this source of heterogeneity may help explain why some individuals always drink alcohol but never smoke cigarettes. It also allows for heterogeneity in the cost of working and schooling caused by unobserved characteristics such as ability. The final line in equation 4 is an idiosyncratic shock to the nonpecuniary utility flow that person i receives at time t from making the choice (hc_{it}, sub_{it}).

3.2 State Variables

The state variables in the vector s_{it} can be divided into continuous state variables s_{it}^c and discrete state variables s_{it}^d. The continuous state variables are the stocks of substance use. The discrete state variables include measures of human capital as well as the number of arrests.

I allow for the accumulation of addictive stocks in each substance, which captures the potential costs of changing substance use from one period to the next. These are the addictive stock associated with cigarette use ($\mathscr{C}_{it}$); with alcohol use ($\mathscr{A}_{it}$); and with marijuana use ($\mathscr{M}_{it}$). The addictive stock associated with cigarettes evolves as follows:

$$\mathscr{C}_{it} = \begin{cases} \beta_1^c \mathscr{C}_{i,t-1} & if\ cigs_{i,t-1} = 1 \\ \mathscr{C}_{i,t-1} + \beta_2^c & if\ cigs_{i,t-1} = 2 \\ \mathscr{C}_{i,t-1} + 1 & if\ cigs_{i,t-1} = 3 \end{cases}. \tag{6}$$

If an individual chose not to consume cigarettes last period, $cigs_{i,t-1} = 1$, then his addictive stock will depreciate by $\beta_1^c \in (0, 1)$. If instead an individual chose to consume a moderate amount, $cigs_{i,t-1} = 2$, then his addictive stock will increase by $\beta_2^c \in (0, 1)$. $\mathscr{A}_{it}$ and $\mathscr{M}_{it}$ evolve in the same way. β_1^c and β_2^c are important, because they allow me to capture certain patterns of use, such as it being easier for an individual to not use a substance in the current period if he did not use the substance in the previous period. Given this specification, the continuous state variables are $s_{it}^c = (\mathscr{C}_{it}, \mathscr{A}_{it}, \mathscr{M}_{it})$.

Human capital is measured by educational attainment and work experience. Educational attainment is summarized by the years of schooling G_{it} an individual has completed, where

$$G_{it} = G_{i,t-1} + g_{it}. \tag{7}$$

$g_{it} = 1$ if an individual completes a year of schooling and zero otherwise. A student who enrolls in school will not necessarily pass that grade. If an individual enrolls in school, the probability that he completes the grade, π_{it}^{school} , is exogenously determined by the individual's states and choices. That is

$$\pi_{it}^{school} = \Phi\left(\eta_{school}\left(sub_{it}, s_{it}\right) + \mu_i^{pass}\right), \tag{8}$$

where, $\Phi(\cdot)$ is a standard normal cdf, η_{school} is the deterministic part of the probability, and μ_i^{pass} represents the individual's unobserved ability in school.

Work experience is measured by total hours H_{it} worked at any firm and by firm-specific human capital τ_{it}. I only keep track of the firm-specific human capital at the most recent in order to simplify the state space. H_{it} as

$$H_{it} = \begin{cases} H_{i,t-1} & if\ h_{it} = 0 \\ H_{i,t-1} & with\ prob\ 1-\pi^H\ if\ h_{it} = 1 \\ H_{i,t-1}+1 & with\ prob\ \pi^H\ if\ h_{it} = 1 \\ H_{i,t-1}+1 & if\ h_{it} = 2 \end{cases} \tag{9}$$

where π^H is the probability that a part-time job increases human capital by one unit. This probability is estimated along with the other parameters of the model.[5] τ_{it} evolve as

$$\tau_{it} = \begin{cases} 0 & if\ h_{it} = 0\ or\ ne_{it} = 1 \\ \tau_{i,t-1}+1 & if\ h_{it} > 0\ and\ ne_{it} = 0 \end{cases}. \tag{10}$$

So, tenure goes up by one if an individual decides to work at the same firm he worked at last year; that is, $h_{it} > 0$ and $ne_{it} = 0$. Tenure goes to zero if he chooses not to work or to work at a new firm.

Lastly, if an individual gets arrested in period t, then $r_{it} = 1$. Therefore, the number of arrests evolves as

$$R_{it} = R_{i,t-1} + r_{it}. \tag{11}$$

The deterministic part of the probability of getting arrested, η_{arrest}, is a function of substance use, previous arrest history, age, and local law enforcement characteristics. The probability is also a function of an individual's unobserved propensity to get arrested, μ_i^{arrest}. The probability of arrest is written as

$$\pi_{it}^{arrest} = \Phi\left(\eta_{arrest}\left(sub_{it}, s_{it}\right) + \mu_i^{arrest}\right) \tag{12}$$

The discrete state variables are $s_{it}^d = (G_{it}, g_{it-1}, enroll_{i,t-1}, R_{it}, H_{it}, \tau_{it}, h_{i,t-1})$.

3.3 The Individual Optimization Problem

I assume individuals are forward-looking. Each period an individual makes choices in order to maximize his present discounted value of expected lifetime utility subject to the budget constraint in Equation 2. At the beginning of the first period, the individual knows the wage function of each hours choice and the deterministic components of the utility function. He also knows his skill endowment μ_i^w and his choice-specific preference values $\mu_i^N = \left(\mu_i^{sub}, \mu_i^u, \mu_i^P, \mu_i^F, \mu_i^s, \mu_i^{arrest}, \mu_i^{pass}\right)$.

[5] This is a simplification used to decrease the size of my state space.

Let $\mu_i = \left(\mu_i^w, \mu_i^N\right)$. Lastly, the individual knows his current levels of the addictive stocks $\mathscr{C}_{i0}$, $\mathscr{A}_{i0}$, and $\mathscr{M}_{i0}$. Future realizations of firm-specific match values ψ_{ik}, wage shocks ν_{ikt}^P and ν_{ikt}^F, and the choice-specific utility shocks corresponding to the human capital accumulation choices ϵ_{it}^{hc}, and to the substance use choices ϵ_{it}^{sub}, are unknown to the individual until they occur; however, he knows the distributions of these variables. He also knows the probabilities that he will get arrested and will complete a year of schooling conditional on enrollment.

The optimization problem can be represented in terms of choice-specific value functions which give the lifetime discounted value of each choice for a given set of state variables, s_{it}.

The value function for individual i in period t is

$$
\begin{aligned}
V\left(s_{it}\right) \quad &= \max_{hc_{it} \in K_{it}(s_{it}), sub_{it}} v_{it}\left(hc_{it}, sub_{it} | s_{it}\right) \\
s.t. \quad &\quad CONS_{it} \geq C_{min}
\end{aligned}
\tag{13}
$$

where the choice set that individual i faces, $K_{it}\left(s_{it}\right)$, varies with s_{it} because individuals who worked last period have an additional discrete choice, whether to continue working at the same job. The term $v_{it}\left(hc_{it}, sub_{it} | s_{it}\right)$ in Equation 13 is the choice-specific value function

$$
v_{it}\left(hc_{it}, sub_{it} | s_{it}\right) = u_{it}\left(hc_{it}, sub_{it} | s_{it}\right) + \delta EV\left(s_{i,t+1} | s_{it}, hc_{it}, sub_{it}\right), \tag{14}
$$

where δ is the discount factor. $EV\left(s_{i,t+1} | s_{it}, hc_{it}, sub_{it}\right)$ is the expected value function in period $t+1$ conditional on states and choices made in period t. Expectations are taken over the random shocks to utility and wages, future match values at new jobs, future arrests, and future educational attainment. For an individual who is not enrolled in school,

$$
\begin{aligned}
EV\left(s_{i,t+1} | s_{it}, hc_{it}, sub_{it}\right) \quad &= \quad \pi_{it}^{arrest} EV\left(s_{i,t+1} | s_{it}, hc_{it}, sub_{it}, r_{it} = 1\right) \\
&\quad + \left(1 - \pi_{it}^{arrest}\right) EV\left(s_{i,t+1} | s_{it}, hc_{it}, sub_{it}, r_{it} = 0\right)
\end{aligned}
\tag{15}
$$

where π_{it}^{arrest} is the probability that the individual gets arrested. At time t, the individual does not know if he will get arrested, $r_{it} = 1$, but he knows π_{it}^{arrest}. Therefore, his expectations are taken over the probability of arrest. If someone is enrolled in school, then the expected value function includes the probability that the individual passes that grade, $g_{it} = 1$. That is,

$$
\begin{aligned}
EV\left(s_{i,t+1}|s_{it},hc_{it},sub_{it}\right) &= \pi_{it}^{school}\pi_{it}^{arrest}EV\left(s_{i,t+1}|s_{it},hc_{it},sub_{it},g_{it}=1,r_{it}=1\right)\\
&+\pi_{it}^{school}\left(1-\pi_{it}^{arrest}\right)EV\left(s_{i,t+1}|s_{it},hc_{it},sub_{it},g_{it}=1,r_{it}=0\right)\\
&+\left(1-\pi_{it}^{school}\right)\pi_{it}^{arrest}EV\left(s_{i,t+1}|s_{it},hc_{it},sub_{it},g_{it}=0,r_{it}=1\right)\\
&+\left(1-\pi_{it}^{school}\right)\left(1-\pi_{it}^{arrest}\right)EV\left(s_{i,t+1}|s_{it},hc_{it},sub_{it},g_{it}=0,r_{it}=0\right).
\end{aligned}
$$

If someone works part-time, then the expected value will similarly incorporate the probability that human capital, $H_{i,t+1}$, increases by one. Individuals use the value functions to determine the optimal educational, employment, and substance use choices each period.

3.4 Identification

In this subsection, I will discuss the identification of the causal effects of substance use on outcomes as well as the identification of the parameters in the model. The causal effects of substance use on education and career outcomes are difficult to identify since individuals endogenously select into substance use, enrollment, and employment states. One concern is that, for example, those who are more likely to use large amounts of marijuana are those for whom the labor market would offer lower wages regardless of marijuana use, other things held constant. Alternatively, it may be the case that the only individuals using marijuana are those who will not be negatively affected by it. In this case, ignoring unobservable characteristics that are important in both selection into marijuana use and wages will overstate the negative effects of marijuana use on wages. In order to identify the effects of substance use on outcomes, I need selection into enrollment, substance use, and employment states to each be partly explained by exogenous variation that is not correlated with the outcome of interest. I use the cost of higher education in the respondent's state of residence as an exclusion restriction to help explain educational choices; the prices of substances to help explain substance use; local labor market conditions to help explain employment outcomes; and local law enforcement characteristics to help explain arrests.[6]

Next, I will discuss the identification of the parameters in the model. Individuals are observed making different choices because they differ in characteristics that affect their opportunities, for example through wages. The parameters of the wage equations are identified by the co-variation in wages and observable individual characteristics. I also assume that individuals have unobserved characteristics like ability that affect their wages. Based on observables, I can predict a person's wage. An individual with high unobserved ability will have persistently higher than predicted

[6] I show in the appendix that substance prices affect whether or not 14 year olds start using substances and local unemployment rates affect whether or not an individual is working.

wages. This identifies the distributional parameters of individual heterogeneity in productivity. Some individuals may have higher-than-predicted wages because they have a particularly high worker-firm match value, perhaps because they get along with their boss; if such people change jobs, they may no longer have a higher than predicted wage. The parameters of the distribution of the worker-firm match value are identified by the correlation in the residuals of wages within a job after controlling for persistent personal effects that arise across jobs. The distributional parameters of the wage shock are identified by the variation in wages across individuals conditional on individual characteristics and worker-firm match values, arising because wages suddenly rise, for example, for a worker who appears to have low ability and low match value.

Individuals do not necessarily choose the option that maximizes pecuniary utility. The co-variation in observed individual characteristics and choices that are not explained by maximizing pecuniary utility identify the parameters of the nonpecuniary utility. Individuals who are observationally equivalent make different work, schooling, and substance use decisions due to heterogeneity in time-persistent preferences. Based on observables, I can predict what the optimal choice is for an individual. The correlation in residuals in observed choices within an individual identifies the distributional parameters of the unobserved tastes that affect nonpecuniary utility.

4 Data

4.1 NLSY97

I use the 1997 National Longitudinal Survey of Youth (NLSY97), which consists of 4,599 male youths who were 12 to 16 years old as of December 31, 1996. Interviews have been conducted annually since 1997; I use the first 13 waves of data, until they were aged 24-28 in 2009. The NLSY97 collects information about labor market behavior, educational experiences, and drug use. Since substance use is a sensitive topic, these questions are administered through the use of audio computer-assisted self-interview technology rather than an interviewer, so I treat the answers as truthful.[7] The NLSY97 consists of a nationally representative core sample and a supplement that over-samples blacks and Hispanics. I use the entire sample in my analysis. I limit my sample to individuals whom I observe at age 14, which is the first age at which individuals are asked about their substance use. An individual remains in the data set until the observation is truncated

[7]Harrison at al. (2007) report on a validity study of self-reported substance use. The study was conducted in conjunction with the National Household Survey on Drug Abuse, which uses the same technology as the NLSY97. Urine samples were collected for a subset of respondents in order to compare self-reported substance use with actual substance use. For tobacco, 88.7 percent of individual's self-reported use in the past 3 days agreed with their urine sample; 7.7 percent reported no use and tested positive for use while 3.6 reported using and tested negative. For marijuana, 93 percent of individual's self-reported use in the past 3 days agreed with their urine sample; 5.2 percent reported no use tested positive and 1.8 percent reported using and tested negative.

at the first instance of missing information for any variable that I use. My preliminary sample consists of 1,170 individuals whom I observe on average for 7.65 years, providing 8,946 person-year observations.[8]

The decision period of my model corresponds to a school year, which runs from August to July. I use monthly school enrollment arrays to construct enrollment status. My model allows individuals to fail a grade; therefore, I consider an individual as enrolled in a particular school year if he reports being enrolled for at least 1 month of that school year. Among those enrolling in grades K-12, 10.96 percent of my sample fails to advance to the next grade, whereas 29.87 percent of those enrolled in college fail.[9] The amount of education accumulated is determined using a variable that indicates the highest grade completed as of the interview date.

The NLSY97 has week-by-week reports of an individual's working status. I use weekly hours of work in a school year to determine annual hours. I classify an individual as working full-time if he works at least 30 hours per week and 45 weeks per year, or at least 1,350 hours per year. To determine the relevant wage for each working individual, I sum up the hours worked each week during the school year for each job. The wage rate for that school year is the wage at the job where the individual worked the most hours. Annual earnings is defined as that wage multiplied by 2,000 hours for full-time workers and 1,000 for part-time workers.[10] The NLSY97 offers two measures of hourly wages. The first is the hourly wage where overtime and performance pay is excluded. The second includes all extra compensation such as over-time, tips, and bonuses. I choose the second measure because it gives me a better measure of the income individuals have to spend on consumption. Additionally, in my model, I assume that the wage reflects the marginal productivity of the individual, and it is reasonable to assume that increases in performance pay indicate higher productivity, making the measure that includes tips and bonuses more appropriate.

In each wave, individuals are asked questions about their use of cigarettes, alcohol, and marijuana in the past 30 days. I assume that this information is representative of their use over the entire school year.[11] Individuals in my model choose how many cigarettes to consume, how many days to consume alcohol and how many days to consume marijuana. In particular, I allow individ-

[8]Table 14 in the appendix shows how I obtain my preliminary sample. The biggest decrease in my sample size comes from dropping individuals who are not interviewed at age 14. I am assuming that all individuals have not used substances prior to age 14. Therefore, their addictive stock is zero at age 14. If I do not observe a person starting at age 14, then I do not know what their substance use stock is. At a future date, I plan to adapt my estimation strategy to simulate this missing information, which will allow me to use more individuals. Data augmentation methods provide a simple way to do this within my Bayesian MCMC estimation strategy.

[9]According to Acaldi et al. (2011), the median time to graduation for 2008 bachelor's degree recipients to earn their degree was between 52 and 80 months, depending on their path to graduation.

[10]Restricting the choice of hours worked to either 1,000 or 2,000 hours is done for tractability.

[11]I compare the self-reported substance use rates in the NLSY97 with those in the 1997 National Household Survey on Drug Abuse (NHSDA) in Table 15 in the appendix. Rates are slightly higher in the NLSY97, but overall are comparable to those in the NHSDA.

uals to choose whether to consume none of a substance, a moderate amount, or a heavy amount. I classify moderate cigarette use as consuming more than zero but less than 300 cigarettes per 30 days; moderate alcohol use as more than zero but less than 13 days; and moderate marijuana use as more than zero but less than 11 days. The amount consumed in each category is determined by the mean consumption of that substance observed in the data. Figure 1 shows the distribution of substance use among those using a positive amount as well as by moderate and heavy use.

4.2 Price Data

I use the cigarette prices reported in Volume 46 of The Tax Burden on Tobacco (Orzechowski and Walker 2011).[12] It includes federal taxes, state taxes, and the average retail price for a pack of cigarettes from 1955-2010. Table 1 presents summary statistics of the total price of a pack of cigarettes by year, which I define as the average retail price plus taxes. I show that the average price of cigarettes increases significantly over my time span. It rises from \$2.64 per pack in 1997 to \$4.52 per pack in 2008. In addition, there is significant variation across states.

As is common in the literature, I use the Cost of Living Index collected by the American Chamber of Commerce Research Association (ACCRA) for alcohol prices.[13] The index is reported quarterly and collects prices from around 300 cities for a 6-pack of beer. I assume that individuals living in the same state face the same prices, so I aggregate prices to state-year observations. I first average each city's price across quarters to get a city-year price. Then, to get a state price, I calculate the weighted average where weights are proportional to an included city's population.[14] The Cost of Living Index does not collect data from every city each quarter, so I am still left with some states that do not have a price during certain years. I impute these prices using the regression

$$P_{sy} = \beta_1 + \beta_2 statetax_{sy} + \beta_3 \mathbb{1}\left(year_y > 2000\right) + \alpha_s + \eta_y + \epsilon_{sy},$$

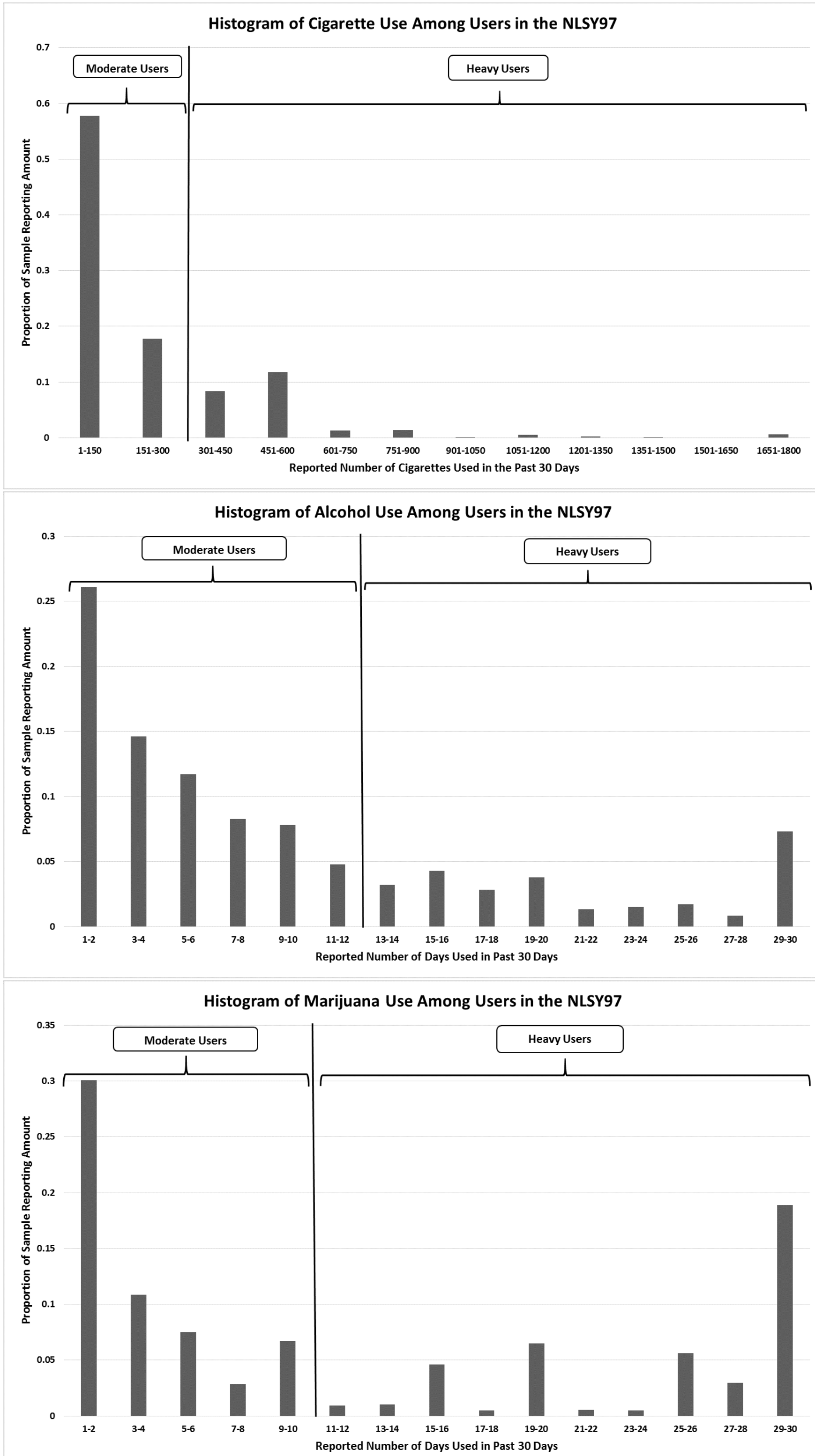
where P_{sy} is the beer price in state s and year y, $statetax_{sy}$ is the state beer tax, α_s is the state fixed effect, and η_y is the year fixed effect. Prior to 2000, ACCRA collected the price of a 6-pack of Budweiser or Miller, and then switched to Heineken in 2000. Therefore, I include $\mathbb{1}\left(year_y > 2000\right)$ as a regressor to adjust for the change in the type of beer sampled after 2000. The regression produces an R-squared of .97. I then impute the missing values using the estimated coefficients and subtract β_3 from observations after the year 2000 to account for the sampling change. Regression results are presented in the appendix. I present summary statistics of the prices I use in Table 1. The

[12] This publication can be found at http://www.taxadmin.org/fta/tobacco/papers/Tax_Burden_2011.pdf.

[13] Ruhm et. al. (2012) find that "barcode" scanner data, collected by AC Nielsen provide a better price elasticity of demand for beer than the ACCRA data. While I would prefer to use scanner data, the data only started being collected in 2006 and only recently became accessible to researchers.

[14] The populations were collected from the U.S. Census Bureau.

Figure 1: Histograms of Substance Use Among Users in the NLSY97



Data are from the NLSY97. The sample used is described in Table 14.

average price of beer decreases over time from \$4.65 for a 6-pack in 1997 to \$4.15 in 2008.

I use marijuana price data assembled from High Times magazine.[15] In each monthly issue, contributors to the "Trans High Market Quotations" (THMQ) section write in with a description of the marijuana, the price per ounce, and the city and state they live in. Using the description of the marijuana, I divide observations into what I am calling low-grade marijuana and high-grade marijuana, where high-grade marijuana has a higher potency of THC, the active drug in marijuana. To divide the observations into grades, I found nicknames for low quality marijuana on marijuana forums and classified an observation as low-grade if it includes keywords such as: Schwag, Brick, Dirt, Mids, Commercial. More details can be found in Alford (2013). I aggregate both the low-grade and high-grade prices to the state-year level in the same way as I do with the alcohol prices. I then estimate the regression,

$$\begin{aligned} P_{sy} &= \beta_1 + \beta_2 medicalmjlaw_{sy} + \beta_3 decriminalized_{sy} + \beta_4 violentcrimerate_{sy} \\ &+\beta_5 murderrate_{sy} + \beta_6 propertycrimerate_{sy} + \beta_7 lowgrade \\ &+\alpha_s + \eta_y + \epsilon_{sy} \end{aligned}$$

in order to impute missing state-year prices, where $medicalmjlaw_{sy}$ equals one if state s has a legalized medical marijuana in year y and $decriminalized_{sy}$ equals one if marijuana is decriminalized in state s. I get an R-squared of .81. Regression results are presented in the appendix. I then impute missing values using the estimated coefficients. I use low-grade prices as the price individuals in my model face because it is unlikely that youth are purchasing and using medical-grade marijuana. Table 1 shows that marijuana prices vary over time and across states, but not in a systematic way like beer and cigarette prices.

For the price of college, I use data collected from several years of the National Tuition and Fee Report published by the Washington Student Achievement Council.[16] The survey collects information on the average undergraduate tuition at over 200 state public institutions. I use the resident undergraduate tuition and required fees for 30 semester credit hours at the state's flagship university as the price of college.

[15] Mireille Jacobson provided me data from 1996-2005, and Mark Anderson provided me with data from 2008-2010. I collected data from 2006-2007 online at www.hightimes.com. See Jacobson (2004) and Anderson, Hansen, and Rees (2013) for more information about the price data. Most papers in the literature use marijuana prices from the Drug Enforcement Administration's (DEA) System to Retrieve Information from Drug Evidence (STRIDE). The DEA's focus is on harder drugs like heroin, so the number of marijuana observations are small, and a large amount of those observations are concentrated in Washington D.C.. The data collected from High Times has much better geographic coverage.

[16] Andrew Barr provided me with this data.

Table 1: Substance Price Summary Statistics, by Year

	Cigarette		Beer		Marijuana	
Year	Mean	Std. Dev.	Mean	Std Dev.	Mean	Std. Dev.
1997	2.64	(0.49)	4.65	(0.39)	123.62	(37.81)
1998	2.78	(0.57)	4.62	(0.37)	113.10	(47.35)
1999	2.97	(0.61)	4.63	(0.37)	134.62	(37.66)
2000	3.72	(0.62)	4.33	(0.39)	110.83	(30.21)
2001	3.81	(0.65)	4.27	(0.41)	124.13	(49.35)
2002	4.04	(0.66)	4.34	(0.44)	115.98	(44.38)
2003	4.38	(0.89)	4.34	(0.44)	123.54	(35.99)
2004	4.41	(0.91)	4.35	(0.44)	118.94	(49.46)
2005	4.39	(0.98)	4.23	(0.34)	108.66	(42.52)
2006	4.45	(1.01)	4.17	(0.37)	103.06	(40.80)
2007	4.41	(1.06)	4.24	(0.41)	124.05	(38.19)
2008	4.52	(1.09)	4.15	(0.42)	105.89	(30.43)

Data on cigarette prices come from volume 46 of The Tax Burden on Tobacco and are for one pack of cigarettes. A pack of cigarettes generally contains 20 cigarettes. Data on beer prices come from ACCRA's Cost of Living Index and are for a 6-pack of beer. Data on marijuana prices come from High Times Magazine and are for one ounce of Marijuana. I use the CPI to adjust all prices to 2000 dollars.

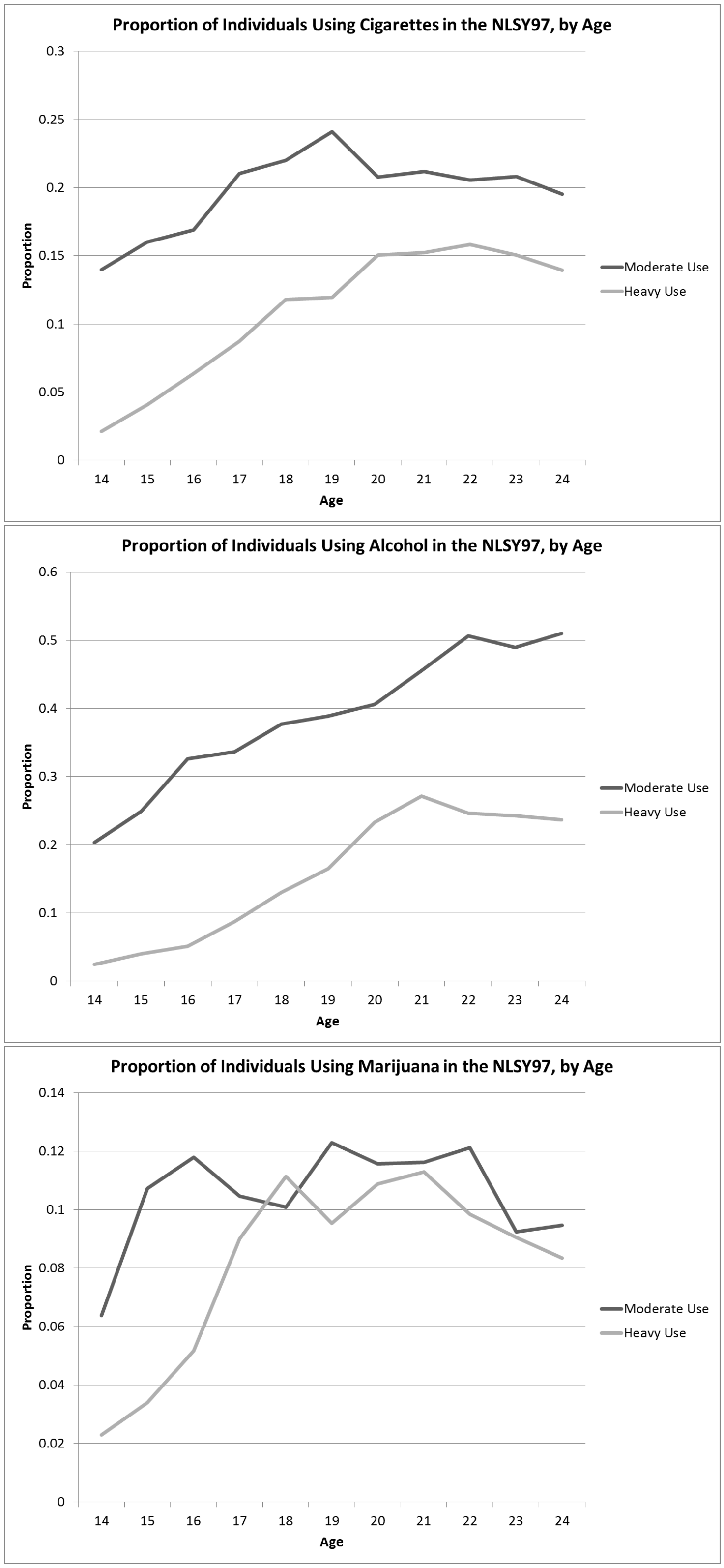
4.3 Descriptive Statistics

This subsection discusses some key characteristics about substance use among individuals in my sample. Figure 2 shows the proportion of individuals using each substance by age. I classify individuals as using either none of a substance, a moderate amount, or a heavy amount. Details are presented in Figure 1. In general, substance use increases as individuals get older. The proportion of individuals using moderate amounts of cigarettes increases until age 19, where 24 percent of my sample reports using a moderate amount of cigarettes. The proportion using heavy amounts of cigarettes increases to its peak of around 15 percent at age 20. The proportion drinking moderately increases until it flattens at age 22, where 50 percent of my sample reports drinking moderately. On the other hand, the proportion reporting drinking heavily hits its peak at age 21 at 27 percent. The percentage of individuals using moderate amounts of marijuana is around 12 percent for almost my entire sample, whereas those using heavy amounts increases drastically from 2.3 percent at age 14 to 10.9 percent at age 21.

I run several regressions to see how substance prices are associated with use at age 14. Table 2 presents the estimates and the marginal effects of an ordered probit of substance use on prices.[17] The dependent variable is the level of use at age 14: no use, moderate use, or heavy use. I assume

[17] I do not present all of the marginal effects in this section. Please contact the author if you would like to see additional marginal effects.

Figure 2: Proportion Using Substances in the NLSY97, by Age



Data come from the NLSY97. The sample used is described in Table 14.

that individuals have not used substances prior to age 14, so these regressions measure how prices affect the initiation of substance use. Marginal effects predict the probability of an individual not using. Table 2 shows that a one dollar increase in the price of a pack of cigarettes is associated with

Table 2: Ordered Probit: Substance Use at Age 14

	Cigarette		Alcohol		Marijuana	
	Estimates	Marginal Effects	Estimates	Marginal Effects	Estimates	Marginal Effects
Cigarette Price	-0.187**	0.045**	0.167**	-0.049**	0.233**	-0.036**
	(0.09)	(0.02)	(0.08)	(0.02)	(0.10)	(0.02)
Beer Price	0.074	-0.018	-0.141	0.041	-0.234*	0.036*
	(0.10)	(0.02)	(0.09)	(0.03)	(0.12)	(0.02)
Marijuana Price	-0.001	0.000	-0.003***	0.001***	-0.002	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Black	-0.499***	0.119***	-0.429***	0.126***	-0.063	0.010
	(0.12)	(0.03)	(0.11)	(0.03)	(0.13)	(0.02)
Hispanic	-0.190*	0.046*	-0.042	0.012	-0.097	0.015
	(0.11)	(0.03)	(0.10)	(0.03)	(0.13)	(0.02)
N	1,175		1,175		1,175	

Marginal effects predict the probability of an individual not using. Cigarette price reflects the price of a pack of cigarettes. A pack generally contains 20 cigarettes. Beer price reflects the price of a 6-pack of beer. Marijuana price reflect the price of an ounce of marijuana. A joint of marijuana contains around half of a gram.
* p<0.10, ** p<0.05, *** p<0.01

a 4.5 percentage point increase in the probability that an individual does not use cigarettes at the age of 14. The price of beer is not statistically significantly associated with whether an individual drinks, but I show that a dollar increase in the price of a pack of cigarettes is associated with a 4.9 percentage point decrease in the probability of not drinking at age 14; a one dollar increase in the price of low-grade marijuana is associated with a 0.1 percentage point increase in the probability of not drinking. Lastly, I show that a dollar increase in cigarette prices is associated with a 3.6 percentage point decrease in the probability of not using marijuana; a one dollar increase in the price of beer is associated with a 3.6 percentage point increase in the probability of not using. The price of marijuana is not statistically significantly associated with the individuals marijuana choice at age 14.

Next, I run several regressions to see how past substance use is associated with current substance use, work, and education outcomes at age 23. Results are presented in the appendix. None of the relationships I describe should be interpreted as causal, but they do describe key patterns in the data. The main concern is that the enjoyment that an individual gets from substance use is likely correlated with the educational and employment outcomes of interest for other reasons. The measures of past substance use that I use in these regressions are the total number of years the individual chose to consume any amount of a substance. For all substances, past substance use increases the probability of using that substance at age 23. However, past cigarette, alcohol, or marijuana use does not affect the probability of using the other substances. I find that past cigarette use decreases educational attainment at age 23, but past alcohol and marijuana use increase it. This counterintuitive result may arise because, for example, those who use marijuana and do not get arrested also have some sort of unobserved characteristic that is correlated with higher education. Past cigarette use decreases the probability of working while past marijuana use increases. Past cigarette and marijuana use do not statistically significantly affect the wages of those working at age 24; past alcohol use increases them.

5 Estimation Strategy

I estimate the parameters of my model using Bayesian Markov Chain Monte Carlo (MCMC) methods. Using classical estimation techniques to estimate my model is difficult for several reasons. First, it is important in my model that I allow for a substantial amount of time-persistent unobserved heterogeneity in order to identify the causal effect of state dependence on outcomes. It is generally burdensome to incorporate unobserved heterogeneity using classical methods because these terms must be integrated out in order to calculate the objective function, yet they generally do not have closed form solutions. MCMC estimation offers the convenience of data augmentation that allows me to avoid integration when evaluating my value function. Second, estimation with classical methods often require evaluating the value function for each sample observation and for each trial guess of the parameters. This can be very time consuming in problems like mine with large choice sets and large state spaces. In particular, my model suffers from the curse of dimensionality due to the three continuous state variables pertaining to substance use stocks. Imai, Jain, and Ching (2009) (IJC) develop an estimator that approximates the value functions by using stored value functions from earlier iterations of the MCMC algorithm, making estimation feasible.[18] Lastly, my likelihood function is highly non-linear and probably not globally concave. This can make finding a global maximum difficult. The MCMC method is theoretically guaranteed to converge to the posterior distribution, which can be used to calculate the global maximum of the likelihood function.

5.1 Econometric Specification

Before discussing the estimation of my model in more detail, I specify the distributions of the random variables as well as the functional forms of the utility flow equations.

[18] Keane and Wolpin (1994) and Rust (1997) discuss ways to break the curse of dimensionality. The method developed in Rust (1997) does not apply to problems where the continuous state variable is deterministic, such as mine. Neither paper addresses the issue of allowing substantial unobserved heterogeneity.

5.1.1 Distributional Assumptions

I assume that permanent worker-firm productivity match values and true randomness in wages are distributed as follows:

$$\begin{aligned}\psi_{ik} &\sim iidN\left(0,\sigma_{\psi}^{2}\right)\\ \nu_{itk}^{h} &\sim iidN\left(0,\sigma_{h}^{2}\right).\end{aligned}$$

Let $\psi_i = \{\psi_{ik}\}_{k=1}^{K_i}$ denote the vector of all job match offers an individual receives in his lifetime. Since an individual receives a new full-time and part-time job offer each period, the number of job matches an individual receives over his lifetime is $K_i = 2 \cdot T$.

I assume that the random shocks to the substance use decisions ϵ_{it}^{sub} are independent across individuals and time. I assume the random shocks to the human capital choices ϵ_{it}^{hc} are independent across individuals, time, and discrete choices. Specifically, $\epsilon_{it}^{sub} \sim iidN\left(0,\Sigma_{sub}\right)$ and $\epsilon_{it}^{hc} \sim iid\ EV$. As is true in all models with discrete choices, the parameters in my model are identified only up to a scale, so I set the variance of ϵ_{it}^{hc} equal to $\frac{\pi^2}{6}$.

I assume the parental transfer shocks are distributed log-normally and are independent across individuals, time, and discrete choices. I let the mean and standard deviation vary by the discrete hours and enrollment decisions; that is,

$$\ln\left(\tilde{w}_{it}\left(hc_{it}\right)\right) \sim iidN\left(\mu_{\tilde{w}}\left(hc_{it}\right),\sigma_{\tilde{w}}^{2}\left(hc_{it}\right)\right).$$

Lastly, I cannot separately identify an individual's time-persistent heterogeneity in preferences for nonemployment, part-time work, and full-time work. I set $\mu_i^{h=0}$ to zero, and I assume that the time-persistent individual heterogeneity in productivity μ_i^w and in preferences $\mu_i^N = \left(\mu_i^{sub},\mu_i^P,\mu_i^F,\mu_i^s,\mu_i^{arrest},\mu_i^{pass}\right)$

$$\mu_i = \begin{pmatrix}\mu_i^w\\ \mu_i^N\end{pmatrix} \sim N\left(0,\Sigma_\mu\right)$$

5.1.2 Utility Flow Equations

In this section, I parameterize the deterministic parts of the pecuniary and nonpecuniary utility flows. Table 3 summarizes the variables I include in each equation. Specific functional forms can be found in the appendix. Table 3 additionally clarifies the exclusion restrictions which I discussed in Section 3.4 and which help identify my model.

The elements of the deterministic part of the wage equation $\theta_w\left(sub_{it},s_{it}\right)$ are presented in the first column. Including experience and tenure in the wage equation captures how general and firm-

Table 3: Empirical Specification for Deterministic Parts of the Utility Function

	Pecuniary Utility	Nonpecuniary Utility		
Variables Included In Each Equation	Wage Equation	Working	School	Substance Use
Hours Part-Time	Yes*	Yes*	Yes	
Hours Full-Time			Yes	
Experience	Yes	Yes		
Experience Squared	Yes			
Tenure	Yes	Yes		
Tenure Squared	Yes			
Worked Part-Time Last Period		Yes		
Worked Full-Time Last Period		Yes		
Education	Yes	Yes	Yes	
Enrolled Last Period			Yes	
Passed Last Period			Yes	
Age Less Than 18	Yes			
Age Between 18 and 21	Yes			
Substance Use	Yes	Yes	Yes	Yes
Substance Use Interacted with Other Use				Yes
Substance Use Interacted with Substance States				Yes
Substance Use Interacted with Age				Yes
Unemployment Rate	Yes	Yes		
Arrests	Yes	Yes	Yes	
Constant	Yes	Yes	Yes	Yes

* The wage and nonpecuniary utility are multiplied by the work part-time coefficient when individuals choose to work part-time.

specific human capital affect wages. An indicator for younger ages is included to capture employer preferences about hiring young individuals and individuals who may be enrolled in college. Substance use history is included because it may limit productivity. Arrests are included so that I can separately identify the effect of substance use on wages from the effect of being arrested. Substance use is associated with more arrests; so, if I exclude arrests and arrests decrease productivity, then the effect of substance use on wages may be downward-biased. I include the unemployment rate because the unemployment rate may affect the distribution of wage offers.

The deterministic part of the nonpecuniary utility of employment, $\theta_N \left(sub_{it}, s_{it}\right)$, is presented in the second column. Current substance use is included because there may be a nonpecuniary cost to working if an individual is using a lot of substances. The deterministic part of the nonpecuniary utility from not working, $\theta_N^0 \left(sub_{it}, s_{it}\right)$ is set equal to zero because the coefficients in the utility flow equation are identified only relative to a base choice.

The nonpecuniary cost function of attending school, $\kappa_s \left(h_{it}, sub_{it}, s_{it}\right)$, is presented in column 3. I include whether the individual was enrolled last period in order to capture the cost of going to school after a period of not enrolling. An indicator of whether he passed if he was enrolled is also included; it is likely that an individual who did not pass last period has a higher nonpecuniary cost to enrolling this period. The cost of enrollment is allowed to vary by hours worked. Current substance use is included because an individual using large amounts of alcohol may have higher nonpecuniary costs of enrolling in school than an individual who is not drinking.

The nonpecuniary utility gained from using substances, $\alpha \left(sub_{it}, s_{it}\right)$ is presented in column 4. I include the interaction between substance use and substance states to capture the addictiveness and gateway effects of substances. For example, a positive coefficient on the interaction of today's heavy cigarette use and past cigarette stock suggests that cigarettes are addictive and that there is a higher cost to quitting if an individual has used high amounts of the substance in the past. A positive coefficient on the the interaction of today's marijuana use and past cigarette use suggests that cigarette use is a gateway to marijuana use. Interactions between current substance uses are included to capture patterns such as individuals enjoying to smoke when they drink. These addiction, gateway, and complementarity effects are assumed to affect the utility of substance use but not the utility of employment or education choices conditional on substance use. Lastly, I include an indicator for whether or not an individual is using which captures any fixed costs associated with using a substance.

5.1.3 State Transition Probabilities

The probabilities of completing a grade of school and of getting arrested can be expressed as probit models with latent variables $G^*_{i,t+1}$ and $R^*_{i,t+1}$, respectively. That is,

$$
\begin{aligned}
G^*_{i,t+1} &= \eta_{school}\left(sub_{it}, s_{it}\right) + \mu_i^{pass} + N\left(0,1\right) \\
R^*_{i,t+1} &= \eta_{arrest}\left(sub_{it}, s_{it}\right) + \mu_i^{arrest} + N\left(0,1\right)
\end{aligned}
$$

where these probabilities depend on many of the state variables

$$
\begin{aligned}
\eta_{school}\left(sub_{it}, s_{it}\right) &= \eta_1^s + \eta_2^s G_{it} + \eta_3 G_{it}^2 + \eta_4^s \mathbb{1}\left(G_{it} > 12\right) + \eta_5^s \mathbb{1}\left(G_{it} > 16\right) + \\
&\quad \eta_6^s enroll_{i,t-1} + \eta_7^s g_{i,t-1} enroll_{i,t-1} + \eta_3^s G_{it}^2 \eta_8^s \mathbb{1}\left(cigs_{it} = 2\right) + \\
&\quad \eta_9^s \mathbb{1}\left(cigs_{it} = 3\right) + \eta_{10}^s \mathbb{1}\left(alc_{it} = 2\right) + \eta_{11}^s \mathbb{1}\left(alc_{it} = 3\right) + \\
&\quad \eta_{12}^s \mathbb{1}\left(mj_{it} = 2\right) + \eta_{13}^s \mathbb{1}\left(mj_{it} = 3\right)
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
\eta_{arrest}\left(c_{it}, d_{it}, s_{it}\right) &= \eta_1^a + \eta_2^a age_{it} + \eta_3^a age_{it}^2 + \eta_4^a R_{it} + \eta_5^a X_{it}^R + \\
&\quad \eta_6^a \mathbb{1}\left(cigs_{it} = 2\right) + \eta_7^a \mathbb{1}\left(cigs_{it} = 3\right) + \\
&\quad \eta_8^a \mathbb{1}\left(alc_{it} = 2\right) + \eta_9^a \mathbb{1}\left(alc_{it} = 3\right) + \\
&\quad \eta_{10}^a \mathbb{1}\left(mj_{it} = 2\right) + \eta_{11}^a \mathbb{1}\left(mj_{it} = 3\right) + \\
&\quad \eta_{12}^a \mathbb{1}\left(alc_{it} = 2\right) \mathbb{1}\left(age_{it} \geq 21\right) + \eta_{13}^a \mathbb{1}\left(alc_{it} = 3\right) \mathbb{1}\left(age_{it} \geq 21\right)
\end{aligned} \tag{17}
$$

X_{it}^R are local law enforcement characteristics for individual i, which are not included elsewhere in the model. With this specification, a person completes a grade of school ($g_{it} = 1$) if $G^*_{i,t+1} > 0$ and a person gets arrested ($r_{it} = 1$) if $R^*_{i,t+1} > 0$. Therefore, conditional on $G^*_{i,t+1}$ and $R^*_{i,t+1}$ the transition of the state variables from period t to period $t+1$ is deterministic.

5.2 Estimation Algorithm

As in most dynamic structural estimation algorithms, my problem can be divided into an outer loop and an inner loop. The outer loop estimates the parameters, while the inner loop calculates value functions that are used in the outer loop to calculate the objective function. I use Bayesian estimation in the outer loop to estimate the parameters of the model. Traditional approaches use backwards recursion in the inner loop to solve for the value functions. The IJC algorithm provides a way to approximate the value functions by using stored value functions from earlier iterations of the MCMC algorithm, making estimation more likely to be feasible. The rest of this subsection presents details of the outer and inner loops.

5.2.1 The Outer Loop

In Bayesian estimation, rather than estimating the parameters by optimizing an objective function, the researcher does so by simulating from the posterior distribution of the parameters. The posterior distribution is made up of the likelihood function and any prior beliefs the researcher has about the parameters. By Bayes' rule, the posterior distribution $\Pr(\theta|data) = \frac{\Pr(data|\theta)\Pr(\theta)}{\Pr(data)}$, where $\Pr(data|\theta)$ is the likelihood function, and $\Pr(\theta)$ are the prior beliefs. The approach is to make many draws of the parameters from the posterior distribution so that moments of the posterior can be calculated.

Usually, the posterior distribution does not have a convenient form from which to take draws. In this case, the MCMC method can be used to generate draws from the posterior distribution. To estimate the parameters of my model, I use Gibbs sampling in conjunction with the Metropolis-Hastings (M-H) algorithm and data augmentation.[19] Instead of taking draws from the multi-dimensional posterior for all the parameters, Gibbs sampling allows me to take draws for a subset of parameters called blocks, conditional on the values of the other parameters. The M-H algorithm is utilized in the case that the conditional posterior for a block of parameters does not take a simple form that can be easily drawn from.

In order to make calculation of the likelihood easier, I also use a data augmentation step within my Gibbs sampler. The usefulness of data augmentation was first discussed in Albert and Chib (1993). McCulloch and Rossi (1994) then showed how to use data augmentation within a Gibbs sampler to estimate a static multinomial probit model. Instead of simulating the posterior of the parameters, I simulate the joint posterior distribution of the parameters and latent variables, such as unobserved wage offers and time-persistent unobserved heterogeneity. In the data augmentation block, the latent variables are simulated conditional on the parameters and the data. Then, in the parameter block, the Gibbs sampler simulates the parameters conditional on the data and the latent variables. So, in this step the data has been augmented to include the latent variables, which makes drawing the parameter values easier by eliminating the need to perform high-dimensional integration. After repeating these steps, the resulting sequence of simulated parameters and latent variables is a Markov chain with the stationary distribution equal to the joint posterior distribution of the parameters and the latent variables.

[19] Chapters 9 and 12 in Train (2009) present a good introduction to Bayesian estimation, MCMC estimation, Gibbs sampling, and the M-H algorithm. Geman and Geman (1984), Gelfand and Smith (1990), Casella and George (1992), Smith and Roberts (1993), and Tierney (1994), among many others, discuss M-H and Gibbs sampling.

5.3 The Inner Loop

Within the blocks of the outer loop, I need to calculate the choice-specific value functions in order to calculate the likelihood function; this is the inner loop. As seen in equation 14, the choice-specific value function, to do this I need to be able to calculate the expected value function $EV\left(s_{i,t+1}|s_{it}, hc_{it}, sub_{it}\right)$. Following the IJC method, I approximate the expected value function by using saved information from earlier iterations of the MCMC estimation algorithm.[20] At the end of each MCMC iteration, I randomly pick one person and I calculate and store pseudo value functions for that person at each point in the discrete state space. To approximate the expected value function, I use a weighted average of the stored pseudo value functions, where an observation gets a higher weight if the current parameter values are close to the parameters that were used to calculate the stored value function.

To be more specific, at the end of MCMC iteration r, I have stored vectors of pseudo value functions $\left\{\left\{\tilde{V}^{l}_{i't}\left(s^{d}_{t}, s^{cl}_{t}|\theta^{*l}_{i'}\right)\right\}_{s^{d}_{t}\in S^{d}_{t}}\right\}_{t=1}^{T}$ for Q people, where i' is a randomly selected individual whose pseudo-value function was stored at the end of the lth iteration, $\theta^{*l}_{i'}$ are the parameters and latent variables from the lth iteration of the MCMC algorithm, and s^{cl}_{t} is a random draw of the three continuous state variables drawn from an uniform distribution over the support of the continuous state variables. The pseudo-expected value function is

$$\widetilde{EV}\left(s_{i,t+1}|s_{it}, hc_{it}, sub_{it}\right) = \sum_{l=r-Q}^{r-1} \tilde{V}^{l}_{i't}\left(s^{d}_{t}, s^{cl}_{t}|\theta^{*l}_{i'}\right) \cdot K_{H_{\theta}}\left(\theta^{*l}_{i'} - \theta^{cr}_{i}\right) \cdot K_{H_{s}}\left(s^{cl}_{t} - s^{cr}_{t}\right), \qquad (18)$$

where K_H is a Gaussian Kernel with bandwidth matrix H. This is simply a weighted average of the stored pseudo-value function where the weights are determined by how close the parameter variables, latent variables, and continuous state variables are to the current values.

It is important to note that random draws of the idiosyncratic shocks are used in the calculation of the stored pseudo value functions $\tilde{V}^{l}_{i't}\left(s^{d}_{t}, s^{cl}_{t}|\theta^{*l}_{i'}\right)$. However, these variables are not included in $\theta^{*l}_{i'}$, because these values are unknown to the agent at time period $T-1$. This method allows those variables to be integrated out. Lastly, this method allows me to use kernel-based local interpolation me to overcome the curse of dimensionality caused by the three continuous state variables. As seen in Equation 18, I am weighing how close the continuous state variables are to previous values. Doing this allows me to calculate the pseudo value functions at only one randomly drawn vector of the continuous state space per parameter update.

Further details about my estimation strategy can be found in the appendix.

[20]See Ching et. al (2012) for a more detailed description of the method. Also, see Osborne (2011) and Zhou (2011) for applications of the algorithm.

6 Results

In this section, I present preliminary estimates of the structural parameters of my model. While the current simplifications are relatively minor, I discuss later the steps I am taking towards estimating the complete version of my model.[21] These estimates were obtained from estimating my model on a limited sample of 150 white males, 150 black males, and 150 Hispanic males. I currently simplify the model in several ways. First, I set the weight individuals put on pecuniary utility α equal to 0.5, so they value pecuniary and nonpecuniary utility the same. I assume that the probability that a part-time job increases human capital by one unit π^H equals one. Additionally, I assume the random shocks to the substance use decisions are $\epsilon_{it}^{sub} \sim iidN(0,1)$, rather than estimating Σ_{sub}. Lastly, I assume that the minimum level of consumption $C_{min} = \$11,000$. If an individual were to make a choice that would put their consumption below C_{min}, the I provide them with enough income so that that $CONS_{it} = C_{min}$. Results are presented in Tables 4 through 13. I ran the MCMC estimation algorithm for approximately 30,000 iterations. I used the 20,000th to the 30,000th iterations to derive the posterior means and standard errors, which are presented below. The standard errors of the parameters are simply the standard deviations of the posterior distribution. There are too many parameters to discuss each one individually, so instead I focus my discussion on key parameter estimates and what the estimates reveal about the substance use and human capital accumulation decision process.

6.1 Nonpecuniary Utility Flows

The estimated parameters of the nonpecuniary utility flows are presented in Tables 4 through 6. These parameter estimates represent the effect of each variable on the nonpecuniary utility of enrolling in school, working part-time, working full-time, and using substances. All estimates are relative to the base choice of not enrolling, not working, and not using substances.

I first present the estimates of the nonpecuniary utility of enrolling in school in Table 4. The results show that an extra year of education decreases the utility of enrolling in school by 0.631 log-yearly consumption units for white males (0.459 for Hispanic males and 1.184 for black males). The decrease in the schooling utility flow is equivalent to the effect of a 46.8 percent (35.8 percent and 69.4 percent) decrease in yearly consumption. Working part-time raises the nonpecuniary utility of attending school (because it has a positive coefficient) and so is a complement to enrollment, whereas working full-time is a substitute. Consuming a heavy amount of cigarettes has a negative sign and so is a substitute for enrolling in school for Hispanic and black males. For these individuals, heavy cigarette use is equivalent to decreasing yearly consumption by 90.2 percent

[21] Relaxing the simplifications is not difficult. Most of these simplifications were made in order to speed up estimation. With the simplifications, estimation takes approximately 72 hours versus 240 hours with out them.

and 95.2 percent, respectively. Heavy use of alcohol has a large negative effect on the utility of schooling for black males, but not for white or Hispanic males. Consuming heavy amounts of marijuana decreases the utility of enrolling for all males.

Table 4: Non-Pecuniary Utility of Attending School Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Education	-0.631*	(0.02)	-0.459*	(0.05)	-1.184*	(0.09)
Enrolled Last Period	13.306*	(0.33)	12.757*	(0.35)	17.327*	(0.70)
Passed Last Period	11.646*	(0.43)	11.513*	(0.54)	14.293*	(0.46)
Working Part-time	0.615*	(0.30)	2.035*	(0.25)	0.103	(0.42)
Working Full-time	-4.200*	(0.24)	-3.567*	(0.50)	-5.568*	(0.44)
Arrests	-0.981*	(0.14)	-0.487*	(0.27)	-0.934*	(0.14)
Moderate Cigarettes	0.515*	(0.26)	-0.680*	(0.27)	-0.284	(0.39)
Heavy Cigarettes	-0.328	(0.23)	-2.322*	(0.52)	-3.032*	(0.73)
Moderate Alcohol	0.276*	(0.14)	0.555*	(0.23)	0.328	(0.28)
Heavy Alcohol	-0.222	(0.41)	1.705*	(0.26)	-7.661*	(0.90)
Moderate Marijuana	1.967*	(0.22)	-0.120	(0.18)	-0.262	(0.29)
Heavy Marijuana	-1.983*	(0.65)	-3.935*	(0.56)	-7.925*	(1.49)
Constant	-0.794*	(0.20)	-3.598*	(0.41)	-2.778*	(0.38)

* $p<0.05$

I present the estimates of the nonpecuniary utility of working in Table 5. Individuals get higher nonpecuniary utility from working part-time than they do from working full-time. An extra year of experience or tenure increases the utility of working, whereas an extra year of education decreases it. Individuals who worked in the previous period have a higher utility of working in the current period. The coefficients on heavy substance use are all positive, suggesting that heavy substance use and working are complements; a possible explanation for this is that work provides income to pay for substances. Moderate cigarette and marijuana use increase the nonpecuniary utility of working for white males, whereas moderate marijuana use decreases it for black males.

Lastly, I present the estimates of the nonpecuniary utility associated with using substances in Table 6. The coefficients on the interaction of substance use and the addictive stock being equal to zero show there is a large nonpecuniary cost of starting to use a substance for the first time. In particular, there is a significant and large nonpecuniary cost to using a heavy amount of a substance if you have never used that substance before. For example, using a heavy amount of marijuana for the first time is equivalent to a 99.99 percent decrease in yearly consumption. Increasing the addictive stock of a substance increases the utility of using a heavy amount of that substance, suggesting that all substances are addictive. Past cigarette use increases the utility of consuming heavy amounts of alcohol, suggesting that cigarettes are a gateway drug for alcohol. On the other hand, past cigarette use decreases the utility of consuming marijuana. Past alcohol use increases the utility of using moderate amounts of marijuana for black males and increases the utility of using heavy amounts of marijuana for Hispanic males, suggesting that alcohol is a

Table 5: Non-Pecuniary Utility of Working Equation Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Experience	0.220*	(0.01)	0.394*	(0.01)	0.274*	(0.02)
Education	-0.360*	(0.00)	-0.280*	(0.01)	-0.389*	(0.01)
Arrest	-0.925*	(0.05)	-0.865*	(0.05)	-1.915*	(0.06)
Tenure	0.188*	(0.01)	0.073*	(0.02)	0.171*	(0.03)
Worked Part-time Last Period	2.713*	(0.08)	0.816*	(0.15)	2.955*	(0.06)
Worked Full-time Last Period	5.349*	(0.05)	6.391*	(0.08)	7.009*	(0.08)
Unemployment Rate	-0.438*	(0.01)	-0.695*	(0.02)	-0.707*	(0.01)
Moderate Cigarettes	0.807*	(0.05)	-0.089	(0.19)	0.193	(0.15)
Heavy Cigarettes	2.246*	(0.08)	4.517*	(0.27)	4.015*	(0.31)
Moderate Alcohol	0.069	(0.05)	0.157	(0.15)	0.811*	(0.18)
Heavy Alcohol	4.859*	(0.10)	5.268*	(0.09)	4.932*	(0.09)
Moderate Marijuana	0.514*	(0.10)	0.258	(0.20)	-0.801*	(0.14)
Heavy Marijuana	8.441*	(0.15)	7.955*	(0.22)	7.577*	(0.30)
Work Part-Time***	1.088*	(0.05)	1.047*	(0.04)	1.039*	(0.07)
Work Part-time Constant	3.262*	(0.03)	2.624*	(0.11)	3.735*	(0.14)
Work Full-time Constant	0.320*	(0.04)	-1.455*	(0.10)	-0.634*	(0.10)

*** The utility of working is multiplied by the work part-time coefficient when individuals choose to work part-time.
* $p<0.05$

gateway drug among minority males. I additionally find evidence supporting the reverse gateway theory, because I find that past marijuana use increases the utility of using any level of cigarettes. Lastly, the positive coefficients on the interactions of substances show that cigarette, alcohol, and marijuana use are complements. This suggests that policies aimed at reducing the use of one substance, will decrease the use of all three substances. Krauss et al. (2014) also find that cigarette and alcohol are complements and suggest increasing the tax on cigarettes in order to lower alcohol consumption.

6.2 Wage Equation

Estimates of the wage offer equation parameters are presented in Table 7. White, Hispanic, and black males working part-time earn hourly wages 4.13, 4.17, and 0.79 percent lower, respectively, than individuals working full-time. Each extra year of education increases the wages of white males by 4.6 percent, but has no statistically significant effect on black and Hispanic workers. This is likely due to the fact that I only observe individuals until they are 24. The benefits of more education may not have had enough time to show up as higher wages. Each additional arrest decreases the wages of black and Hispanic males by approximately 10 percent, but has no effect on white males. Overall, substance use has little effect on wages. Moderate cigarette use decreases wages for black males and heavy cigarette use decreases wages for Hispanic males. Heavy marijuana use, on the other hand, increases wages for white males.

Table 6: Non-Pecuniary Utility of Consuming Drugs Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Moderate Cigarette Use						
Constant	-2.082*	(0.10)	-0.522*	(0.15)	-0.672*	(0.30)
Cigarette Addictive Stock=0	-3.466*	(0.17)	-2.566*	(0.20)	-3.117*	(0.38)
Cigarette Addictive Stock	-0.297*	(0.02)	-0.512*	(0.01)	-0.632*	(0.02)
Alcohol Addictive Stock	-0.152*	(0.02)	0.138*	(0.01)	-0.049*	(0.01)
Marijuana Addictive Stock	0.036*	(0.01)	0.225*	(0.04)	0.037*	(0.01)
Heavy Cigarette Use						
Constant	-4.7540*	(0.11)	-7.583*	(0.19)	-6.259*	(0.14)
Cigarette Addictive Stock=0	-11.005*	(0.14)	-11.616*	(0.20)	-9.752*	(0.32)
Cigarette Addictive Stock	1.096*	(0.01)	0.905*	(0.02)	0.920*	(0.04)
Alcohol Addictive Stock	-0.268*	(0.01)	-0.528*	(0.04)	-0.170*	(0.01)
Marijuana Addictive Stock	0.231*	(0.01)	0.131*	(0.02)	0.128*	(0.01)
Moderate Alcohol Use						
Constant	-2.740*	(0.17)	-2.971*	(0.18)	-2.697*	(0.27)
Alcohol Addictive Stock=0	-1.097*	(0.15)	-1.203*	(0.13)	-1.707*	(0.16)
Cigarette Addictive Stock	-0.315*	(0.01)	-0.186*	(0.02)	-0.420*	(0.02)
Alcohol Addictive Stock	-0.146*	(0.01)	-0.051*	(0.03)	0.037	(0.03)
Marijuana Addictive Stock	-0.197*	(0.01)	-0.078*	(0.04)	-0.107*	(0.02)
Legally Using	-2.742*	(0.26)	-1.707*	(0.26)	-4.560*	(0.10)
Heavy Alcohol Use						
Constant	-10.710*	(0.17)	-10.997*	(0.19)	-8.453*	(0.34)
Alcohol Addictive Stock=0	-22.936*	(0.32)	-15.506*	(0.18)	-17.422*	(0.14)
Cigarette Addictive Stock	-0.095*	(0.01)	-0.396*	(0.02)	-0.409*	(0.02)
Alcohol Addictive Stock	1.277*	(0.01)	1.071*	(0.01)	1.075*	(0.02)
Marijuana Addictive Stock	0.112*	(0.01)	-0.388*	(0.02)	-0.261*	(0.04)
Legally Using	-1.058*	(0.10)	-1.788*	(0.12)	-0.168	(0.21)
Moderate Marijuana Use						
Constant	-6.066*	(0.11)	-4.469*	(0.09)	-4.601*	(0.17)
Marijuana Addictive Stock=0	-3.153*	(0.19)	-2.356*	(0.16)	-0.720*	(0.14)
Cigarette Addictive Stock	-0.642*	(0.02)	-0.729*	(0.02)	-0.744*	(0.01)
Alcohol Addictive Stock	-0.158*	(0.02)	-0.056*	(0.02)	0.073*	(0.02)
Marijuana Addictive Stock	0.084*	(0.02)	-0.126*	(0.02)	0.049*	(0.02)
Heavy Marijuana Use						
Constant	-11.985*	(0.20)	-11.906*	(0.30)	-8.431*	(0.35)
Marijuana Addictive Stock=0	-13.235*	(0.28)	-9.055*	(0.28)	-15.297*	(0.32)
Cigarette Addictive Stock	-0.414*	(0.01)	-0.378*	(0.01)	-0.667*	(0.03)
Alcohol Addictive Stock	-0.245*	(0.02)	0.088*	(0.02)	-0.089*	(0.05)
Marijuana Addictive Stock	0.726*	(0.02)	0.423*	(0.03)	0.770*	(0.02)
Using Cigarettes and Alcohol	2.256*	(0.13)	2.260*	(0.22)	1.423*	(0.23)
Using Cigarettes and Marijuana	2.296*	(0.13)	1.527*	(0.16)	1.349*	(0.21)
Using Alcohol and Marijuana	3.129*	(0.19)	3.382*	(0.27)	2.019*	(0.28)
Using Cigarettes* Age	-0.230*	(0.01)	-0.271*	(0.02)	-0.034	(0.04)
Using Alcohol* Age	0.278*	(0.01)	0.308*	(0.01)	0.147*	(0.02)
Using Marijuana* Age	-0.151*	(0.02)	-0.217*	(0.03)	0.166*	(0.01)

* $p<0.05$. Age is measures as true age minus 13.

Table 7: Wage Equation Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Education	0.046*	(0.01)	-0.019	(0.02)	0.013	(0.02)
Experience	0.163*	(0.01)	0.197*	(0.02)	0.086*	(0.04)
Experience Squared	-0.019*	(0.00)	-0.012*	(0.00)	-0.008*	(0.00)
Tenure	-0.178*	(0.02)	-0.079*	(0.03)	-0.136*	(0.03)
Tenure Squared	0.039*	(0.00)	0.031*	(0.00)	0.021*	(0.00)
Arrests	-0.054	(0.07)	-0.105*	(0.06)	-0.098*	(0.06)
Age less than 18	-0.068	(0.07)	0.128	(0.14)	-0.092	(0.35)
Age between 18 and 21	-0.190*	(0.08)	-0.035	(0.13)	-0.108	(0.24)
Unemployment Rate	0.045*	(0.01)	0.016	(0.01)	0.036	(0.03)
Moderate Cigarettes	-0.112	(0.10)	-0.096	(0.09)	-0.208*	(0.10)
Heavy Cigarettes	-0.055	(0.08)	-0.266*	(0.12)	0.074	(0.14)
Moderate Alcohol	-0.030	(0.04)	0.005	(0.08)	0.080	(0.09)
Heavy Alcohol	-0.061	(0.07)	0.029	(0.10)	0.077	(0.19)
Moderate Marijuana	0.047	(0.11)	-0.073	(0.10)	-0.157	(0.19)
Heavy Marijuana	0.132*	(0.08)	-0.142	(0.11)	-0.116	(0.16)
Constant	1.419*	(0.13)	2.172*	(0.14)	1.811*	(0.29)
Work Part-Time***	0.959*	(0.04)	0.958*	(0.06)	0.992*	(0.09)
Error Standard Deviations						
True Randomness in Wages	0.787*	(0.02)	0.794*	(0.03)	0.818*	(0.03)
Firm Match Value	0.568*	(0.041)	0.500*	(0.04)	0.599*	(0.045)

*** The wage is multiplied by the work part-time coefficient when individuals choose to work part-time.
* p<0.05

6.3 State Variable

In this sub-section, I present estimates of the parameters that determine how the state variables evolve. These estimates can be found in Tables 8 through 10. First, I show how the addictive stocks evolve in Table 8. Not using cigarettes for a period decreases the addictive stock of cigarettes by 50.8 percent, 53.5 percent, and 15.9 percent for whites, Hispanics, and blacks, respectively. Not using alcohol for a period has nearly no effect on the addictive stock of alcohol. Not using marijuana for a period decreases the addictive stock of whites, Hispanics, and blacks by 82.2 percent, 52.0 percent, and 21.4 percent. Using a moderate amount of a substance does not increase the addictive stocks of substances.

Table 8: Addictive Stock Development Parameters

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
No Cigarette Use	0.492*	(0.01)	0.465*	(0.01)	0.841*	(0.02)
Moderate Cigarette Use	0.000*	(0.00)	0.000*	(0.00)	0.000*	(0.00)
No Alcohol Use	0.999*	(0.00)	0.999*	(0.00)	0.994*	(0.00)
Moderate Alcohol Use	0.000*	(0.00)	0.000*	(0.00)	0.000*	(0.00)
No Marijuana Use	0.178*	(0.01)	0.480*	(0.11)	0.786*	(0.01)
Moderate Marijuana Use	0.000*	(0.00)	0.000	(0.00)	0.000*	(0.00)

* p<0.05

Next, I present the coefficients of the probability that an individual passes a grade in Table 9 and the probability of arrest in Table 10. After grade 7, an extra year of education decreases the probability of passing a grade. While this results may seem counterintuitive, school gets harder as individuals advance, thus making it less likely that they will successfully advance to the next grade. Individuals who enrolled and passed in the previous period have a higher probability of passing. Heavy substance use decreases the probability of passing among all males. As individuals get older, the probability of arrest decreases. An increase in the number of police per person decreases the probability of arrest for white males, but increases it for black males. Heavy substance use increases the probability of arrest for black males, but has no statistically significant effect for Hispanic males.

6.4 Heterogeneity in Preferences and Skills

Lastly, I present the variance-covariance matrix of the time-persistent unobserved heterogeneity in preferences and skills in Tables 11, 12, and 13. For white males, I find that individuals with higher than average preferences for moderate cigarette use also tend to have higher than average preferences for alcohol and moderate marijuana use, lower preferences for working part-time, lower chances of getting arrested, and lower chances of passing a grade. White males who have higher

Table 9: Probability of Passing if Enrolled Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Education	0.125*	(0.00)	0.118*	(0.00)	0.113*	(0.01)
Education Squared	-0.018*	(0.00)	-0.017*	(0.00)	-0.022*	(0.00)
College	0.368*	(0.08)	0.589*	(0.05)	0.380*	(0.04)
Graduate School	-0.087	(0.07)	-0.195*	(0.10)	-0.123*	(0.04)
Enrolled Last Period	0.633*	(0.09)	1.739*	(0.13)	2.046*	(0.05)
Passed Last Period	2.137*	(0.10)	1.862*	(0.05)	3.114*	(0.07)
Moderate Cigarettes	-0.006	(0.04)	0.069*	(0.03)	-0.263*	(0.04)
Heavy Cigarettes	-0.306*	(0.03)	-0.307*	(0.03)	-0.535*	(0.04)
Moderate Alcohol	-0.067*	(0.04)	-0.008	(0.03)	-0.028	(0.02)
Heavy Alcohol	-0.315*	(0.07)	-0.082	(0.05)	-1.008*	(0.04)
Moderate Marijuana	0.149*	(0.04)	-0.005	(0.04)	-0.122*	(0.05)
Heavy Marijuana	-0.116	(0.09)	-0.626*	(0.21)	-0.685*	(0.05)
Constant	0.098	(0.09)	-1.595*	(0.10)	-1.987*	(0.15)

* p<0.05

Table 10: Probability of Arrest Coefficients

	White		Hispanic		Black	
	mean	se	mean	se	mean	se
Age	-0.157*	(0.00)	0.060*	(0.01)	-0.044*	(0.01)
Age Squared	-0.024*	(0.00)	-0.022*	(0.00)	-0.023*	(0.00)
Arrests	2.068*	(0.03)	0.487*	(0.09)	-0.885*	(0.06)
Police Per Person	-27.231*	(3.41)	3.325	(15.46)	21.241*	(6.87)
Moderate Cigarettes	-0.115*	(0.04)	-0.154*	(0.08)	-0.177*	(0.07)
Heavy Cigarettes	-0.094	(0.07)	-0.450*	(0.09)	0.304*	(0.05)
Moderate Alcohol	-0.053	(0.04)	-0.044	(0.04)	-0.005	(0.11)
Heavy Alcohol	-0.167*	(0.08)	0.034	(0.14)	0.816*	(0.11)
Moderate Marijuana	-0.206*	(0.04)	-0.255*	(0.08)	-0.124*	(0.05)
Heavy Marijuana	0.015	(0.14)	0.126	(0.14)	0.796*	(0.26)
Legally Moderate Alcohol	-0.073*	(0.04)	0.068	(0.06)	0.092*	(0.05)
Legally Heavy Alcohol	0.119*	(0.04)	0.045	(0.12)	-0.306*	(0.14)
Constant	-1.292*	(0.15)	-0.966*	(0.16)	-1.503*	(0.08)

* p<0.05. Age is measures as true age minus 13.

preferences for heavy alcohol have higher preferences for moderate marijuana, but lower preferences for heavy marijuana; have stronger preferences for working full-time; have lower probability of arrest; and have lower probability of passing. Having a high preference for heavy marijuana use is positively correlated with a high preference for working, getting arrested, and passing and negatively correlated with earnings.

For Hispanic males, having a high preference for heavy cigarette use is correlated with a high preference for heavy marijuana use, working full-time, and enrolling in school and a lower probability of arrest. Individuals with strong preferences for heavy alcohol use have lower preferences for working, but have higher wages. Those with higher preferences for using heavy marijuana have higher preferences for working part-time and lower preferences for enrolling in school.

Black males with higher than average preferences for heavy cigarettes are likely to have lower preferences for alcohol use and heavy marijuana use, lower preferences for working part-time, higher preferences for enrolling in school and higher probability of passing, higher earnings potential, and higher probability of arrest. Having a higher preference for heavy alcohol use is positively correlated with having higher preferences for marijuana use and working and is negatively correlated with the probability of arrest and the probability of passing. Black males with higher preferences for using heavy marijuana have higher preferences for working part-time, lower preferences for enrolling in school, lower chances of passing, and lower chances of getting arrested.

Table 11: Variance-Covariance of Unobserved Heterogeneity for White Males

	Moderate Cigarettes	Heavy Cigarettes	Moderate Alcohol	Heavy Alcohol	Moderate Marijuana	Heavy Marijuana	Work Part-Time	Work Full-Time	School	Wage	Arrest	Pass
Moderate Cigarettes	1.371* (0.19)											
Heavy Cigarettes	0.111 (0.09)	0.570* (0.08)										
Moderate Alcohol	0.314* (0.09)	-0.001 (0.07)	0.605* (0.08)									
Heavy Alcohol	0.387* (0.12)	-0.019 (0.07)	0.590* (0.10)	1.082* (0.17)								
Moderate Marijuana	1.249* (0.26)	0.244 (0.15)	1.026* (0.17)	1.603* (0.26)	4.000* (0.58)							
Heavy Marijuana	-0.448* (0.12)	-0.127* (0.07)	-0.505* (0.08)	-0.722* (0.12)	-1.630* (0.24)	1.031* (0.13)						
Work Part-time	-0.594* (0.17)	-0.005 (0.10)	-0.670* (0.13)	-1.198* (0.20)	-2.498* (0.40)	1.167* (0.18)	2.016* (0.31)					
Work Full-time	0.019 (0.06)	-0.030 (0.03)	-0.032 (0.03)	0.089* (0.05)	-0.087 (0.09)	0.119* (0.05)	-0.067 (0.07)	0.226* (0.03)				
School	0.040 (0.08)	-0.228 (0.06)	-0.134* (0.05)	0.022 (0.07)	-0.133 (0.16)	-0.017 (0.08)	-0.187* (0.10)	0.028 (0.04)	0.490* (0.07)			
Wage	0.019 (0.06)	0.008 (0.04)	0.039 (0.04)	-0.011 (0.06)	-0.013 (0.10)	-0.110* (0.06)	0.143* (0.08)	-0.189* (0.03)	0.002 (0.04)	0.272* (0.04)		
Arrest	-0.766* (0.24)	0.108 (0.15)	-1.010* (0.17)	-1.582* (0.29)	-3.666* (0.59)	1.457* (0.25)	2.647* (0.43)	-0.059 (0.10)	-0.012 (0.14)	0.238* (0.12)	4.214* (0.65)	
Pass	-0.452* (0.15)	0.037 (0.11)	-0.747* (0.12)	-1.037* (0.18)	-2.302* (0.34)	0.885* (0.15)	1.589* (0.25)	-0.071 (0.07)	0.223* (0.10)	0.151* (0.07)	2.640* (0.38)	1.905* (0.26)

* $p<0.05$

6.5 Simulations

In order to interpret the results presented above, I run the following counter-factual simulation to see how exogenous substance use at age 15 affects outcomes at age 24. I first make 1,000

Table 12: Variance-Covariance of Unobserved Heterogeneity for Hispanic Males

	Moderate Cigarettes	Heavy Cigarettes	Moderate Alcohol	Heavy Alcohol	Moderate Marijuana	Heavy Marijuana	Work Part-Time	Work Full-Time	School	Wage	Arrest	Pass
Moderate Cigarettes	0.781* (0.10)											
Heavy Cigarettes	0.342* (0.09)	0.949* (0.14)										
Moderate Alcohol	0.433* (0.09)	0.571* (0.12)	1.146* (0.16)									
Heavy Alcohol	-0.206* (0.09)	-0.470* (0.11)	0.007 (0.11)	1.000* (0.13)								
Moderate Marijuana	0.758* (0.17)	0.052 (0.16)	0.102 (0.16)	0.584* (0.17)	2.178* (0.44)							
Heavy Marijuana	-0.030 (0.09)	0.174* (0.09)	-0.104 (0.10)	-0.473* (0.10)	-0.668* (0.21)	0.640* (0.12)						
Work Part-time	-0.480* (0.13)	-0.198 (0.12)	-0.325* (0.14)	-0.194 (0.15)	-1.146* (0.35)	0.541* (0.17)	1.237* (0.32)					
Work Full-time	0.258* (0.05)	0.345* (0.07)	0.237* (0.06)	-0.179* (0.08)	0.259* (0.11)	0.071 (0.06)	-0.324* (0.10)	0.309* (0.04)				
School	-0.107* (0.06)	0.124* (0.07)	-0.171* (0.07)	0.063 (0.09)	0.186 (0.16)	-0.130* (0.07)	-0.313* (0.14)	0.114* (0.06)	0.525* (0.09)			
Wage	-0.357* (0.06)	-0.077 (0.07)	-0.065 (0.06)	0.161* (0.07)	-0.425* (0.12)	0.020 (0.06)	0.327* (0.11)	-0.219* (0.04)	-0.101* (0.06)	0.409* (0.06)		
Arrest	-0.533* (0.11)	-0.221* (0.12)	-0.160 (0.11)	-0.092 (0.15)	-1.120* (0.30)	0.193 (0.13)	0.974* (0.27)	-0.419* (0.09)	-0.278* (0.13)	0.400* (0.09)	1.203* (0.27)	
Pass	-0.446* (0.08)	-0.037 (0.09)	-0.108 (0.09)	-0.039 (0.09)	-0.645* (0.17)	0.020 (0.08)	0.483* (0.14)	-0.207* (0.05)	0.068 (0.06)	0.205* (0.06)	0.758* (0.15)	0.731* (0.10)

* p<0.05

Table 13: Variance-Covariance of Unobserved Heterogeneity for Black Males

	Moderate Cigarettes	Heavy Cigarettes	Moderate Alcohol	Heavy Alcohol	Moderate Marijuana	Heavy Marijuana	Work Part-Time	Work Full-Time	School	Wage	Arrest	Pass
Moderate Cigarettes	0.828* (0.17)											
Heavy Cigarettes	-0.845* (0.16)	1.528* (0.22)										
Moderate Alcohol	0.417* (0.13)	-0.429* (0.16)	0.792* (0.11)									
Heavy Alcohol	0.336* (0.08)	-0.428* (0.11)	0.084 (0.08)	0.630* (0.10)								
Moderate Marijuana	-0.305* (0.17)	0.920* (0.28)	-0.212 (0.16)	0.225* (0.10)	1.721* (0.38)							
Heavy Marijuana	1.177* (0.36)	-1.414* (0.40)	0.267 (0.24)	0.380* (0.16)	-1.459* (0.42)	3.652* (0.91)						
Work Part-time	0.649* (0.22)	-0.850* (0.25)	0.360* (0.15)	0.255* (0.09)	-0.818* (0.21)	1.513* (0.43)	1.031* (0.19)					
Work Full-time	-0.047 (0.06)	0.103 (0.09)	0.108 (0.07)	0.141* (0.06)	0.136 (0.09)	-0.357* (0.17)	0.090 (0.09)	0.405* (0.06)				
School	-0.959* (0.25)	0.986* (0.27)	-0.464* (0.12)	-0.158 (0.14)	0.682* (0.31)	-2.338* (0.73)	-0.878* (0.33)	0.519* (0.10)	2.153* (0.47)			
Wage	-0.052 (0.05)	0.113* (0.06)	-0.111* (0.05)	0.054 (0.04)	0.024 (0.06)	0.118 (0.13)	-0.083 (0.05)	-0.062 (0.04)	-0.022 (0.13)	0.278* (0.04)		
Arrest	-1.057* (0.22)	1.035* (0.20)	-0.419* (0.16)	-0.564* (0.12)	0.289 (0.27)	-1.910* (0.62)	-0.783* (0.35)	0.143 (0.09)	1.480* (0.40)	-0.084 (0.11)	2.012* (0.34)	
Pass	-2.624* (0.62)	3.062* (0.66)	-0.808* (0.36)	-1.043* (0.33)	2.025* (0.85)	-6.839* (1.96)	-2.575* (0.95)	1.141* (0.26)	5.266* (1.33)	-0.267 (0.33)	4.580* (1.03)	15.108* (3.57)

* p<0.05

draws from the unobserved heterogeneity distribution. Each of these draws can be thought of as an individual person. I assume that each 15 year-old person enrolls in the 9th grade and passes, and has never worked. I then run three different simulations. In the first simulation, I assume that each individual has never used any substance prior to age 16, that is the addictive stocks at age 16 are equal to zero. In the second simulation, I assume that each individual consumed a moderate amount of each of the substances at age 15. In the third, I assume that each individual consumed a heavy amount of each of the substances at age 15. Using the estimated model, I then simulate the choices of each of these individuals through age 24 and compare the outcomes. In other words, I am forcing 15 year olds in my sample to consume substances and then calculating how the substance use affects outcomes at age 24. Results are presented in Figure 3.

First, I will discuss how exogenous substance use at age 15 affects use at age 24. I find that using a positive amount of substances increases the probability that an individual will be a heavy substance user at age 24. For example, white males who exogenously use a heavy amount of each substance at age 15 are over three times more likely to use a heavy amount of alcohol at age 24 than those that used no substances at age 15. Similarly, Hispanic and black males who use a heavy amount at age 15 are 4.8 and 1.3 times more likely to use a heavy amount of alcohol, respectively. Exogenous use at age 15 also increases the probability that an individual is a heavy cigarette and marijuana user at age 15.

Next, I will discuss how exogenous substance use at age 15 affects educational attainment and wages at age 24. Moderate use at age 15 has no statistically significant effect on the high school graduation rate at age 24. However, heavy use decreases the high school graduation rate of white males and Hispanic males by 5.62 and 7.59 percent, respectively. Dee and Evans (2003) find that teen drinking does not have an affect on educational attainment. The authors do not control for different levels of use in their paper. This result is consistent with my finding that moderate use has no effect on high school graduation rates, but heavy use does. On the other hand, Yamada, Kendix, and Yamada (1996) find that individuals who drink at least two days per week have a 4.58 percent lower chance of graduating high school and that individuals who frequently use marijuana have a 6.28 percent lower probability of graduation. These results are very similar to the effect I find of heavy use on high school graduation rates of white and Hispanic males. Yamada, Kendix, and Yamada (1996) treat substance use as exogenous after controlling for many of observed family and background characteristics. My results provide evidence that this is a reasonable assumption for white and Hispanic males, but perhaps not for black males. My results suggest that unobserved heterogeneity plays an important role for educational attainment of black males because black males who are using heavy amounts of substances will perform badly in school regardless of their substance use.

Although substance use decreases educational attainment for white and Hispanic males, use has no statistically significant effect on wages at age 24, consistent with the estimated wage effects I discussed above. This is likey due to the fact that I only observe individuals until they are 24, so I am unable to obtain good estimates of the return to schooling. In cross-section analysis, several papers find a positive association between past drug use and wages, although these papers treat substance use as exogenous.[22] After controlling for endogenous substance use and for unobserved preferences, I find that past use has no effect on wages. Bray (2005) and Mezza (2011), on the other hand, treat substance use, educational attainment, and work experience as endogenous. Bray (2005) finds that alcohol use while in school or working has a positive effect on human capital accumulation, but that using heavily decreases this gain. Mezza (2011) finds that non-drug users have higher wages than marijuana users. My results may be consistent with those found in Bray (2005) and Mezza (2011) since I am calculating the combined effect of using cigarettes, alcohol, and marijuana.

7 Model Fit

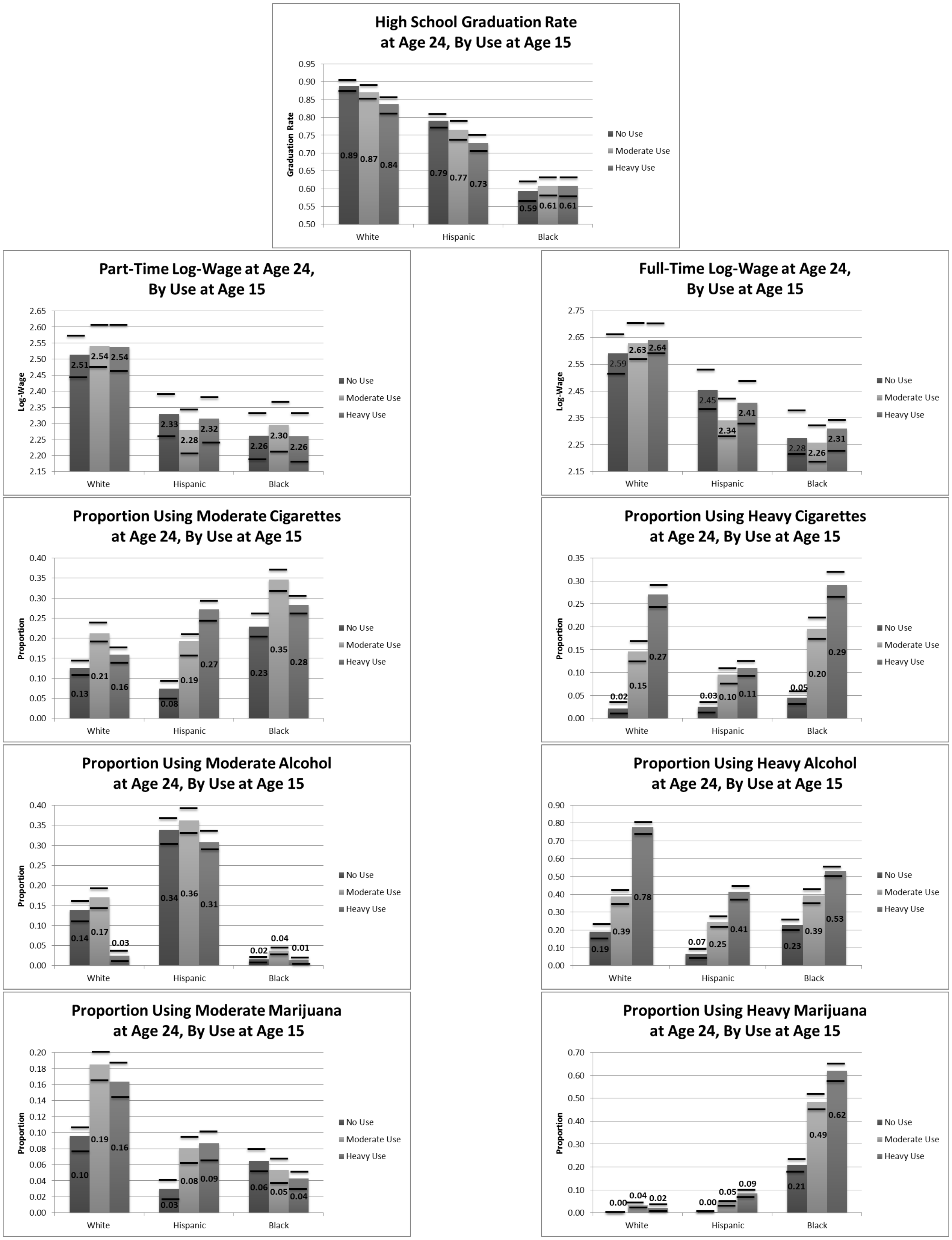
This section examines how well the structural model fits the data. To do this, I simulate educational, employment, and substance use choices using the estimated parameters. I compare these simulated choices with the choices observed in the data. Additionally, I compare the simulated probability of passing a grade and getting arrested with the proportion observed in the data. Lastly, I compare simulated log-wages of those who choose to work generated by the structural model with those observed in the data. Figures 4, 5, and 6 present these comparisons.

Simulated data from the structural model differ from the actual data, although the model fits the overall patterns of enrollment, working, and substance choices quite well. I am able to capture the overall decreasing pattern of enrollment by age. However, I overpredict enrollment for white males, particularly during the ages that they would enroll in college and graduate school, ages 18-21 and 21 and over, respectively. On the other hand, I predict the proportion of Hispanic males enrolling in school quite well. I underpredict the proportion of black males enrolling in school from ages 14 to 17 and overpredict enrollment for ages 18 and 19. I do not include an indicator for college or graduate school in the equation of nonpecuniary utility of school. Including these variables may help match enrollment patterns. I am able to capture the pattern of passing rates quite well for whites and Hispanics, but vastly underpredict the passing rate for black males between the ages of 14 and 17. Rather than having G_{it} and G_{it}^2 in the probability of passing equation, including years in high school and years in college may be able to better capture the observed pattern of passing.

I overpredict the proportion of individuals working part-time and underpredict the proportion

[22]Zarkin et al. (1998), Buchmueller and Zuvekas (1998), among others.

Figure 3: Effect of Substance Use at Age 15 on Outcomes at Age 24



The black lines present the 95th percentile confidence intervals.

of individuals working full-time. This is partly due to the discrepancies between predicted and actual enrollment decisions. Working part-time is a complement for enrolling, whereas working full-time is a substitute. Therefore, when I overpredict enrollment, I will also overpredict part-time work. The simulated log-wages for white males are quite similar to the observed log-wages. I overpredict wages for Hispanic and black males across all ages, but follow the pattern of how wages increase well. Across all races, I do a poor job of fitting the probability of arrest, a pattern that I will investigate further.

Lastly, I discuss the fit of substance use decisions. For alcohol use, I predict a large decrease in moderate alcohol use at the age of 21 for all males. In the current model, I am forcing the time-trend of moderate use and heavy use to be equal. In the data, after the age of 21 many people transition from heavy to moderate use. Therefore, relaxing the assumption that the time-trends are equal may fix this problem. For the most part, I fit the cigarette and marijuana choices well. For black males, I underpredict the amount of heavy cigarettes used after the age of 21, although this is due to the complementarity between alcohol use and cigarette use. I also underpredict the use of heavy marijuana use by Hispanic males after the age of 18.

8 Current Status

I am in the process of estimating the model with my entire sample and relaxing some of the simplifying assumptions. The estimated model will be used for counter-factual simulations to show, for example, the effect of changes in substance prices or in low-skill wages on substance use, education, and employment decisions. These simulations will indicate effective policies to reduce substance use and improve educational and labor market outcomes of young men.

9 Conclusions

In this paper, I develop a dynamic structural model in which individuals make decisions about schooling and work as well as how much alcohol, cigarettes, and marijuana to consume. My model allows substance use to affect career paths through its effects on educational attainment and wage offers. Individuals are rational and forward-looking and make decisions in order to maximize their discounted lifetime utility accounting for the effects of substance use on future outcomes. I improve on the current literature by allowing individuals in my model to make decisions about their level of alcohol, cigarettes, and marijuana use. This allows me to differentiate between the effects of moderate versus heavy substance use. It also allows me to capture the interdependencies of cigarette, alcohol, and marijuana use.

Figure 4: Comparison of Simulated and Actual Choices for White Males

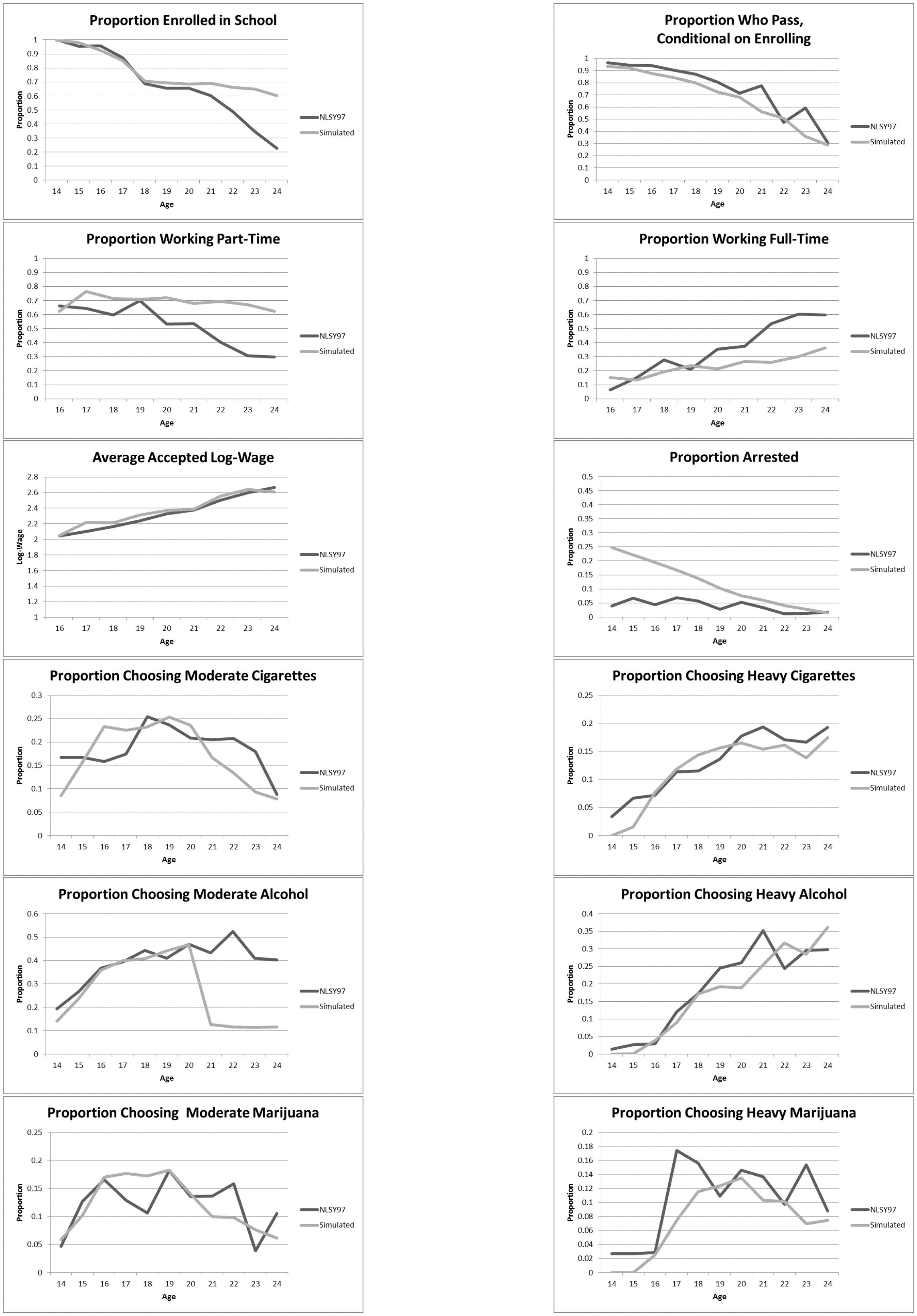


Figure 5: Comparison of Simulated and Actual Choices for Hispanic Males

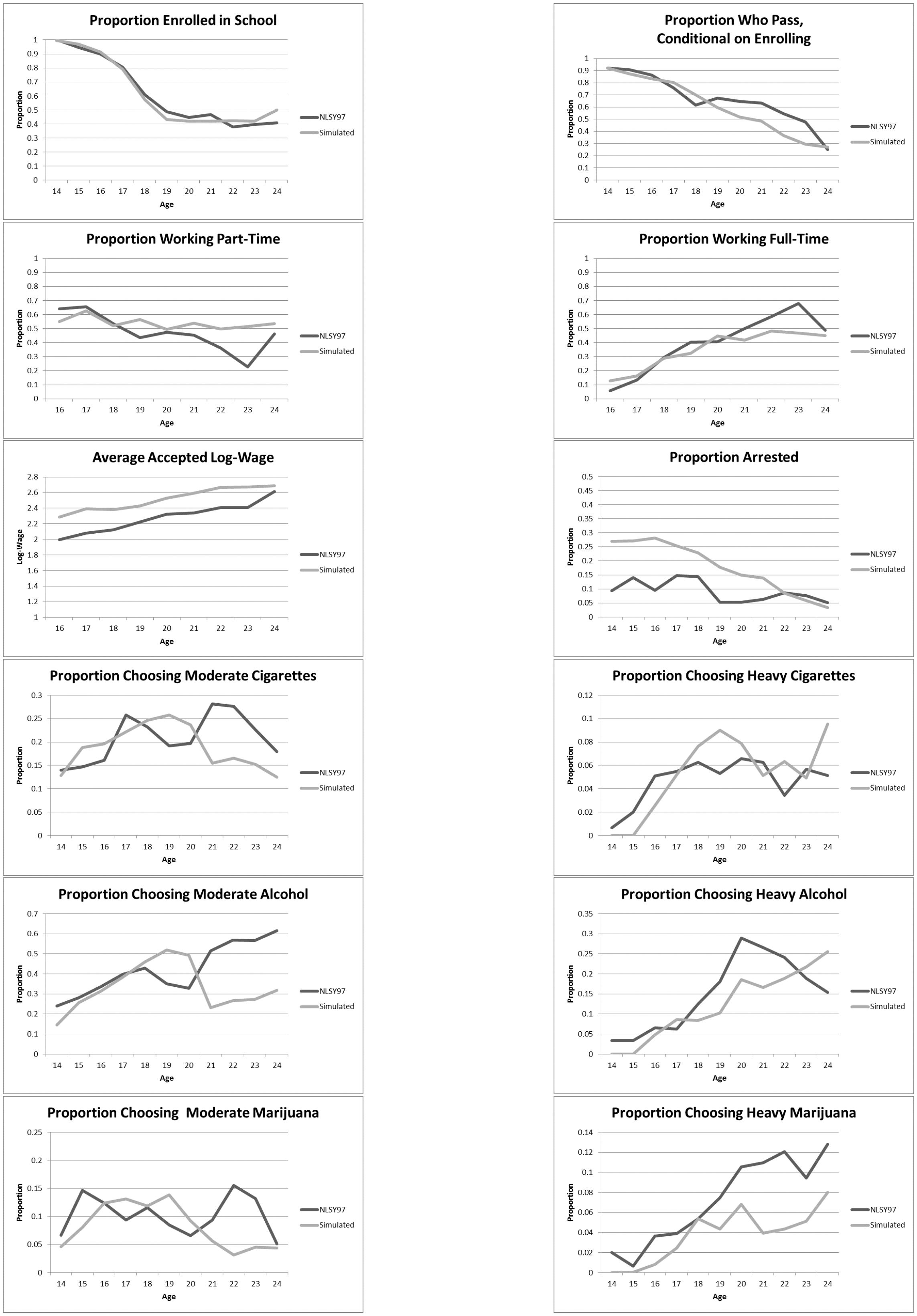
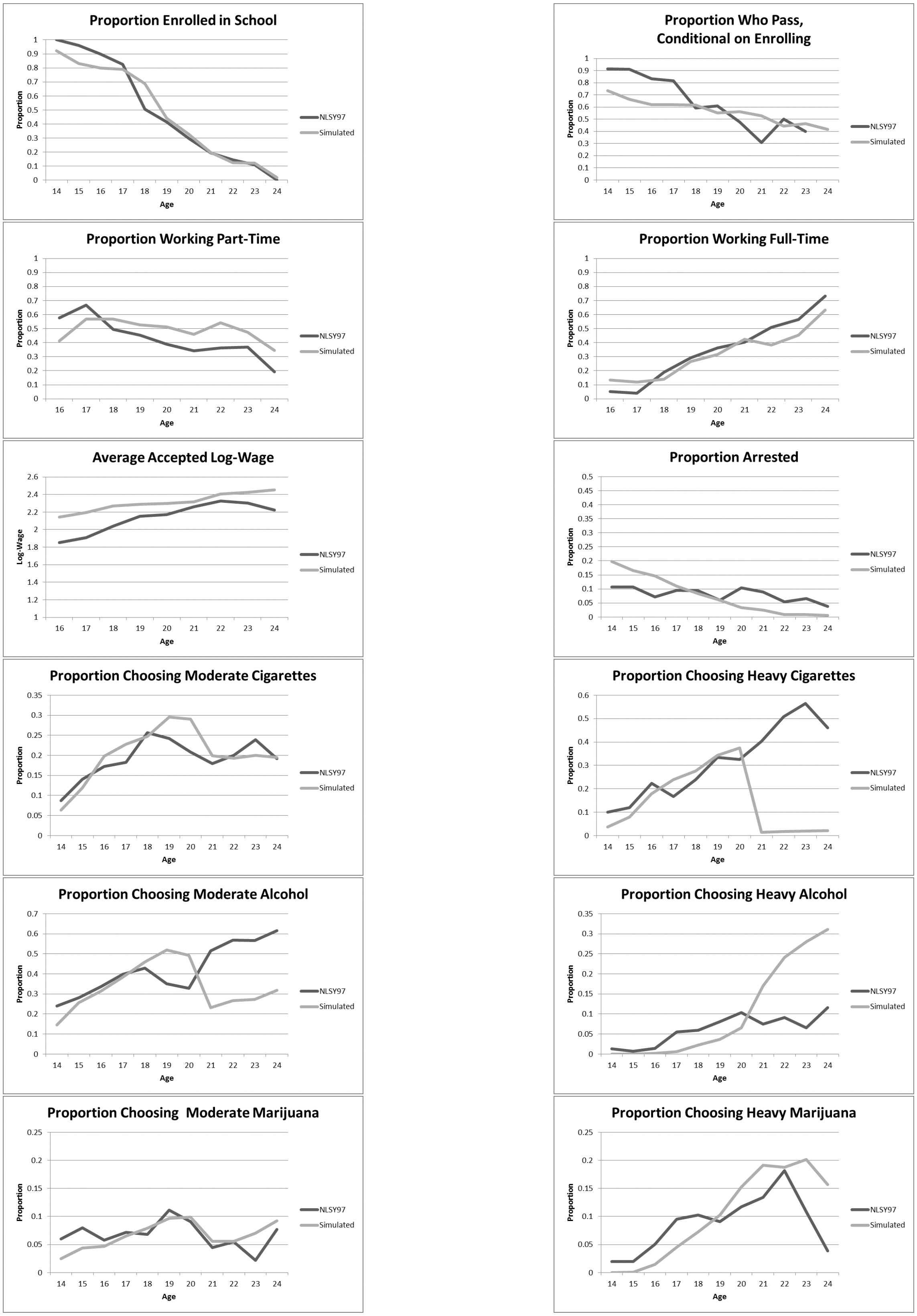


Figure 6: Comparison of Simulated and Actual Choices for Black Males



In preliminary results, I find that exogenous substance use at age 15 substantially increases the probability of being a heavy substance user at age 24. Moderate substance use at age 15 has no effect on high school graduation rates or wages at the age of 24. On the other hand, I find that exogenously using a heavy amount of substances at age 15 decreases the probability of high school graduation of white and Hispanic males by 5.62 and 7.59 percent, respectively, but has no statistically significant effect on log-wage offers at the age of 24. Additionally, I find that cigarette, alcohol, and marijuana use are complements. This suggests that policies aimed at reducing the use of one substance, will decrease the use of all three substances. Lastly, I find that past alcohol use has no direct effect on current marijuana use for white males, but increases the utility of using marijuana for black and Hispanic males. This implies that alcohol is a gateway drug for minority males.

References

[1] James H. Albert and Siddhartha Chib. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of American Statistical Association*, 88(422):669–679, 1993.

[2] Catherine Alford. Medical Marijuana Laws and the Market for Marijuana. 2013.

[3] D Mark Anderson, Benjamin Hansen, and Daniel I Rees. Medical marijuana laws, traffic fatalities, and alcohol consumption. *Journal of Law and Economics*, 56(2):333–369, 2013.

[4] Peter Arcidiacono, Holger Sieg, and Frank Sloan. LIVING RATIONALLY UNDER THE VOLCANO? AN EMPIRICAL ANALYSIS OF HEAVY DRINKING AND SMOKING*. *International Economic Review*, 48(1):37–65, 2007.

[5] Gary S. Becker, Michael Grossman, and Kevin M. Murphy. An Empirical Analysis of Cigarette Addiction. *The American Economic Review*, 84(3):396–418, 1994.

[6] Gary S Becker and Kevin M Murphy. A Theory of Rational Addiction. *Journal of Political Economy*, 96(4):pp. 675–700, 1988.

[7] Jeremy W Bray. Alcohol Use, Human Capital, and Wages. *Journal of Labor Economics*, 23(2):pp. 279–312, 2005.

[8] Jeremy W Bray, Gary A Zarkin, Chris Ringwalt, and Junfeng Qi. The relationship between marijuana initiation and dropping out of high school. *Health Economics*, 9(1):9–18, 2000.

[9] Thomas C Buchmueller and Samuel H Zuvekas. Drug use, drug abuse, and labour market outcomes. *Health Economics*, 7(3):229–245, 1998.

[10] George Casella and Edward I George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[11] E.F. Cataldi, C. Green, R. Henke, T. Lew, J. Woo, B. Sepherd, and P. Siegal. 2008-09 baccalaureate and beyond longitudinal study: First look (nces 2011-236). *U.S. Department of Education. Washington, DC: National Center for Education Statistics.*, 2011.

[12] Frank Chaloupka. Rational Addictive Behavior and Cigarette Smoking. *The Journal of Political Economy*, 99(4):722–742, 1991.

[13] Pinka Chatterji. Illicit drug use and educational attainment. *Health Economics*, 15(5):489–511, 2006.

[14] AndrewT. Ching, Susumu Imai, Masakazu Ishihara, and Neelam Jain. A practitioner's guide to Bayesian estimation of discrete choice dynamic programming models. *Quantitative Marketing and Economics*, 10(2):151–196, 2012.

[15] Phillip J Cook and Michael J Moore. Drinking and schooling. *Journal of Health Economics*, 12(4):411–429, 1993.

[16] Washington Student Achievement Council. National tuition and fee report. Various Volumes, 1996-2010.

[17] Thomas S Dee and William N Evans. Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates. *Journal of Labor Economics*, 21(1):pp. 178–209, 2003.

[18] Jeff DeSimone. Illegal Drug Use and Employment. *Journal of Labor Economics*, 20(4):952–977, 2002.

[19] Zvi Eckstein and Kenneth I Wolpin. Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities. *Econometrica*, 67(6):1295–1339, 1999.

[20] Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

[21] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.

[22] John Geweke. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the ...*, pages 1–14, 1991.

[23] John F Geweke, Michael P Keane, and David E Runkle. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, 80(1):125–165, 1997.

[24] Donna B. Gilleskie and Koleman S. Strumpf. The Behavioral Dynamics of Youth Smoking. *The Journal of Human Resources*, 40(4):822–866, 2005.

[25] Jonathan Gruber and Botond Köszegi. Is Addiction "Rational"? Theory and Evidence. *The Quarterly Journal of Economics*, 116(4):1261–1303, 2001.

[26] Lingxin Hao, V Joseph Hotz, and Ginger Z Jin. Games parents and adolescents play: Risky behaviour, parental reputation and strategic transfers*. *The Economic Journal*, 118(528):515–555, 2008.

[27] Lana D. Harrison, Steven S. Martin, Tihomir Enev, and Deborah Harrington. Comparing drug testing and self-report of drug use among youths and young adults in the general population. Technical report, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD, 2007.

[28] James J Heckman and Bo E Honore. The empirical content of the roy model. *Econometrica: Journal of the Econometric Society*, pages 1121–1149, 1990.

[29] Kosuke Imai and David A van Dyk. A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124(2):311–334, 2005.

[30] Susumu Imai, Neelam Jain, and Andrew Ching. Bayesian Estimation of Dynamic Discrete Choice Models. *Econometrica*, 77(6):1865–1899, 2009.

[31] Mireille Jacobson. Baby booms and drug busts: Trends in youth drug use in the United States, 1975-2000. *Quarterly Journal of Economics*, 119:1481–1512, 2004.

[32] Robert Kaestner. The Effect of Illicit Drug Use on the Labor Supply of Young Adults. *The Journal of Human Resources*, 29(1):pp. 126–155, 1994.

[33] Michael P Keane and Kenneth I Wolpin. The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics*, 76(4):648–672, 1994.

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

[35] Melissa J. Krauss, Patricia A. Cavazos-Rehg, Andrew D. Plunk, Laura J. Bierut, and Richard A. Grucza. Effects of state cigarette excise taxes and smoke-free air policies on state per capita alcohol consumption in the united states, 1980 to 2009. *Alcoholism: Clinical and Experimental Research*, 38(10):2630–2638, 2014.

[36] José M Labeaga. A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics*, 93(1):49–72, 1999.

[37] Ignacio Martinez. A Principal-Agent Model of Parental Transfers and Human Capital Accumulation. 2014.

[38] Robert McCulloch and Peter E Rossi. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240, September 1994.

[39] Alvaro Mezza. Illegal Drugs, Education, and Labor Market Outcomes. 2011.

[40] William Orzechowski and Robert Walker. The tax burden on tobacco. *Historical compilation*, 46, 2011.

[41] Matthew Osborne. Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics*, 9(1):25–70, 2011.

[42] Mark R Rosenzweig and Kenneth I Wolpin. Parental and public transfers to young women and their children. *The American Economic Review*, pages 1195–1212, 1994.

[43] Andrew Roy. Some Thoughts on the Distribution of Earnings. *Oxfor Economic Papers*, 3:135–146, 1951.

[44] Christopher J. Ruhm, Allison Snow Jones, Kerry Anne McGeary, William C. Kerr, Joseph V. Terza, Thomas K. Greenfield, and Ravi S. Pandian. What U.S. data should be used to measure the price elasticity of demand for alcohol? *Journal of Health Economics*, 31(6):851–862, 2012.

[45] John Rust. Using Randomization to Break the Curse of Dimensionality. *Econometrica*, 65(3):487–516, 1997.

[46] Adrian FM Smith and Gareth O Roberts. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–23, 1993.

[47] Paul Sullivan. A DYNAMIC ANALYSIS OF EDUCATIONAL ATTAINMENT, OCCUPATIONAL CHOICES, AND JOB SEARCH*. *International Economic Review*, 51(1):289–317, 2010.

[48] Luke Tierney. Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728, 1994.

[49] High Times. Trans-high market quotations. *New York: Trans-High Corporation Press*, Various Issues, 1996-2010.

[50] Kenneth Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, second edition, 2009.

[51] Tetsuji Yamada, Michael Kendix, and Tadashi Yamada. The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics*, 5(1):77–92, 1996.

[52] Gary A Zarkin, Michael T. French, Thomas Mroz, and Jeremy W Bray. Alcohol use and wages: New results from the national household survey on drug abuse. *Journal of Health Economics*, 17(1):53–68, 1998.

[53] Yiyi Zhou. Bayesian Estimation of a Dynamic Model of Oligopoly Pricing and Entry with Strategic Consumers: Application to Video Games. 2011.

A Data

A.1 Auxiliary Tables

Table 14: Sample Selection

Rule	Individuals Lost
Missing Substance Use or Arrest Information at Age 14	185
Missing Educational Attainment at Age 14	3
Missing Location at Age 14	6
Race Is Mixed	16
Not Interviewed at Age 14	2,270
Only Observed at Age 14	949
Final Sample Size	1,170

Data come from the NLSY97.

The original sample consists of 4,599 male youths who were 12 to 16 years old as of December 31, 1996.

Table 15: Comparison of Substance Use Rates of 12-17 Year-Olds in 1997

	Cigarette		**Beer**		**Marijuana**	
	NLSY97	NHSDA	NLSY97	NHSDA	NLSY97	NHSDA
Total Sample	0.417	0.387	0.450	0.397	0.210	0.189
Female	0.417	0.383	0.442	0.407	0.198	0.182
Male	0.417	0.390	0.458	0.388	0.221	0.195
White	0.445	0.421	0.476	0.425	0.215	0.196
Black, non-Hispanic	0.315	0.282	0.354	0.314	0.183	0.161
Hispanic	0.384	0.330	0.429	0.363	0.208	0.167

This table compares self-reported substance use rates in the 1997 National Longitudinal Survey of Youth with those found in the 1997 National Household Survey on Drug Abuse.

A.2 Descriptive Statistics

I run several regressions to see how past substance use is associated with current substance use, work, and education outcomes at age 23. None of the relationships I describe below should be interpreted as causal, but they do describe key patterns in the data. The main concern is that the enjoyment that an individual gets from substance use is likely correlated with the educational and employment outcomes of interest for other reasons. The measures of past substance use that I use

Table 16: Alcohol Price Imputation Regression

	b/se
State Beer Tax	1.2315***
	(0.21)
Post 2000	2.7718***
	(0.05)
Constant	3.2085***
	(0.23)
N	583
R^2	0.971

The dependent variable is equal to the average price in state s and year t and grade g. I control for state and year fixed effects. I include Post 2000 as a regressor to adjust for the change in the type of beer sampled after 2000. Standard errors are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

Table 17: Marijuana Price Imputation Regression

	b/se
Medical Marijuana Law	-29.4921**
	(13.25)
Medical Marijuana Law and Decriminalized	-2.3898
	(21.81)
Marijuana Decriminalized	11.4161
	(36.82)
Murder Rate	-0.5604
	(2.72)
Property Crime Rate	0.0319***
	(0.01)
Violent Crime Rate	-0.0883*
	(0.05)
Low-Grade Marijuana	-247.0313***
	(4.71)
Constant	266.9717***
	(48.14)
N	887
R^2	0.805

The dependent variable is equal to the average price in state s and year t and grade g. I control for state and year fixed effects. Crime Rates are the rate per 100,000 residents. Standard errors are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

in these regressions are the total number of years the individual chose to consume any amount of a substance.

I first look to see how past substance use affects substance use at age 23. Table 18 presents the results of an ordered probit where the dependent variable is the choice of substance use at age 23; individuals can choose no use, moderate use, or heavy use.[23] As in Table 2, marginal

Table 18: Ordered Probit: Substance Use at Age 23

	Cigarette		Alcohol		Marijuana	
	Estimates	Marginal Effects	Estimates	Marginal Effects	Estimates	Marginal Effects
Cigarette Stock	0.404***	-0.078***	-0.027	0.007	0.024	-0.004
	(0.03)	(0.00)	(0.02)	(0.01)	(0.03)	(0.01)
Alcohol Use Stock	0.002	0.000	0.262***	-0.068***	0.046	-0.008
	(0.03)	(0.00)	(0.03)	(0.01)	(0.04)	(0.01)
Marijuana Use Stock	0.007	-0.001	0.042	-0.011	0.325***	-0.055***
	(0.03)	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)
Cigarette Price	-0.003	0.001	-0.081*	0.021*	-0.039	0.007
	(0.06)	(0.01)	(0.05)	(0.01)	(0.07)	(0.01)
Beer Price	-0.246*	0.048*	-0.117	0.030	-0.275	0.046
	(0.15)	(0.03)	(0.11)	(0.03)	(0.17)	(0.03)
Marijuana Price	0.003**	-0.001**	0.000	0.000	0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Black	0.134	-0.026	-0.107	0.028	0.530**	-0.089**
	(0.18)	(0.04)	(0.14)	(0.04)	(0.22)	(0.04)
Hispanic	-0.285	0.055	-0.160	0.042	0.442**	-0.074**
	(0.18)	(0.03)	(0.14)	(0.04)	(0.19)	(0.03)
N	519		519		519	

Marginal effects predict the probability of an individual not using. Data come from the NLSY97. The sample is limited to those who were interviewed from ages 14 to 23. Cigarette price reflects the price of a pack of cigarettes. A pack generally contains 20 cigarettes. Beer price reflects the price of a 6-pack of beer. Marijuana price reflect the price of an ounce of marijuana. A joint of marijuana contains around half of a gram.
* p<0.10, ** p<0.05, *** p<0.01

effects predict the probability of an individual not using. For all substances, past use increases the probability of using that substance at age 23. However, past cigarette, alcohol, or marijuana use does not affect the probability of using the other substances.

Next, I see how past substance use affects the highest grade attained by age 23. To do this, I run a regression of highest grade attained at age 23 on past substance use and the cost of college. Results are presented in Table 19. Here, there are two regressions: one that excludes arrests and one that includes arrests. In the first regression, past cigarette use decreases educational attainment, while alcohol use increases it. Next, I add in arrests as a covariate and find that the past marijuana use increases the highest grade attained by age 23 after controlling for arrests. This counterintuitive result may arise because those who use marijuana and do not get arrested also have some sort of

[23] I do not present all of the marginal effects in this section. Please contact the author if you would like to see additional marginal effects.

Table 19: Highest Grade Completed by Age 23

	Estimates	Estimates
Cigarette Stock	-0.404***	-0.348***
	(0.04)	(0.04)
Alcohol Use Stock	0.252***	0.251***
	(0.04)	(0.04)
Marijuana Use Stock	0.061	0.105**
	(0.05)	(0.05)
Tuition Price	0.024	0.033
	(0.05)	(0.05)
Black	-0.922***	-0.827***
	(0.25)	(0.24)
Hispanic	-0.824***	-0.777***
	(0.23)	(0.23)
Total Arrests		-0.562***
		(0.09)
N	517	517

Data comes from the NLSY97. The sample is limited to those who were interviewed from ages 14 to 23. Tuition is in 1,000 of dollars.
* p<0.10, ** p<0.05, *** p<0.01

unobserved characteristic that is correlated with higher education.

Next, Table 20 reports ordered probit estimates of how substance use affects full-time, part-time, and no work decisions. The marginal effects I present predict the probability of not working. Past cigarette use decreases the probability that an individual works. After controlling for arrests, marijuana use increases the probability that an individual works by 0.8 percentage points. Past alcohol use does not statistically significantly affect whether an individual works. Additionally, total arrests are not associated with the work decision.

Lastly, I present how past substance use affects the log-wage for individuals who are observed working. Results are presented in Table 21. Cigarette and marijuana use are not statistically significantly associated with wages. However, having used alcohol for one more year increases wages by 4 percent. This regression does not control for selection into working. As suggested in the previous regression, substance use may affect whether or not a person works. Individuals using high level of substances may either be getting low wage offers because of their use or may just not like working. My structural model is able to control for selection, which allows me to better estimate the effect of substance use on wages.

Table 20: Ordered Probit: Work Hours for Individuals at Age 23

	Estimates	Marginal Effects	Estimates	Marginal Effects
Work Experience	0.125***	-0.021***	0.125***	-0.021***
	(0.02)	(0.00)	(0.02)	(0.00)
Education	0.538**	0.007**	0.482**	0.010**
	(0.21)	(0.00)	(0.22)	(0.00)
Education Squared	-0.021***		-0.020**	
	(0.01)		(0.01)	
Cigarette Stock	-0.066***	0.011***	-0.061**	0.010**
	(0.02)	(0.00)	(0.02)	(0.00)
Alcohol Use Stock	-0.010	0.002	-0.006	0.001
	(0.02)	(0.00)	(0.02)	(0.00)
Marijuana Use Stock	0.038	-0.007	0.050*	-0.008*
	(0.03)	(0.00)	(0.03)	(0.00)
Unemployment Rate	-0.118***	0.020***	-0.117**	0.020**
	(0.05)	(0.01)	(0.05)	(0.01)
Black	-0.124	0.021	-0.125	0.021
	(0.14)	(0.02)	(0.14)	(0.02)
Hispanic	0.180	-0.031	0.177	-0.030
	(0.13)	(0.02)	(0.13)	(0.02)
Total Arrests			-0.087	0.016
			(0.11)	(0.02)
Total Arrests Squared			-0.010	
			(0.02)	
N	517		517	

Marginal effects predict the probability of an individual not working. Data come from the NLSY97. Data comes from the NLSY97. The sample is limited to those who were interviewed from ages 14 to 23.
* p<0.10, ** p<0.05, *** p<0.01

Table 21: Log-Wage of Working Individuals at Age 23

	Estimates	Estimates
Experience	0.061*	0.057*
	(0.03)	(0.03)
Experience Squared	-0.004	-0.004
	(0.00)	(0.00)
Education	0.043***	0.037***
	(0.01)	(0.01)
Currently Enrolled in School	-0.085*	-0.082*
	(0.05)	(0.05)
Cigarette Stock	-0.013	-0.010
	(0.01)	(0.01)
Alcohol Use Stock	0.039***	0.040***
	(0.01)	(0.01)
Marijuana Use Stock	-0.016	-0.011
	(0.01)	(0.01)
Total Arrests		-0.059***
		(0.02)
N	484	484

Data comes from the NLSY97. The sample is limited to those who were interviewed from ages 14 to 23.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

B Functional Form of Utility Function

In this section I present the functional forms of the components of the utility function. In the following equations, $\mathbb{1}(\cdot)$ is an indicator that equals one if the argument is true. I assume that the deterministic part of the log-wage equation

$$
\begin{aligned}
\theta_w\left(sub_{it}, s_{it}\right) = {} & \left(\theta_1^w\right)^{\mathbb{1}(h_{it}=1)}\left(\theta_2^w + \theta_3^w G_{it} + \theta_4^w H_{it} + \theta_5^w H_{it}^2 + \theta_6^w \tau_{it} + \theta_7^w \tau_{it}^2\right. \\
& +\theta_8^w \mathbb{1}\left(age_{it} \leq 17\right) + \theta_9^w \mathbb{1}\left(18 \leq age_{it} \leq 21\right) + \theta_{10}^w R_{it} + \theta_{11}^w unempr_{it} \\
& +\theta_{12}^w \mathbb{1}\left(cigs_{it} = 2\right) + \theta_{13}^w \mathbb{1}\left(cigs_{it} = 3\right) \\
& +\theta_{14}^w \mathbb{1}\left(alc_{it} = 2\right) + \theta_{15}^w \mathbb{1}\left(alc_{it} = 3\right) \\
& \left.+\theta_{16}^w \mathbb{1}\left(mj_{it} = 2\right) + \theta_{17}^w \mathbb{1}\left(mj_{it} = 3\right)\right).
\end{aligned}
$$

I assume that the nonpecuniary utility of employment

$$
\begin{aligned}
\theta_n\left(sub_{it}, s_{it}\right) = {} & \theta_1^n \mathbb{1}\left(h_{it} = 1\right) + \theta_2^n \mathbb{1}\left(h_{it} = 2\right) \\
& + \left(\theta_3^n\right)^{\mathbb{1}(h_{it}=1)}\left(\theta_4^n H_{it} + \theta_5^n G_{it} + \theta_6^n \tau_{it} + \theta_7^n R_{it}\right. \\
& +\theta_8^n \mathbb{1}\left(h_{i,t-1} = 1\right) + \theta_9^n \mathbb{1}\left(h_{i,t-1} = 2\right) \\
& +\theta_{10}^n \mathbb{1}\left(cigs_{it} = 2\right) + \theta_{11}^n \mathbb{1}\left(cigs_{it} = 3\right) \\
& +\theta_{12}^n \mathbb{1}\left(alc_{it} = 2\right) + \theta_{13}^n \mathbb{1}\left(alc_{it} = 3\right) \\
& \left.+\theta_{14}^n \mathbb{1}\left(mj_{it} = 2\right) + \theta_{15}^n \mathbb{1}\left(mj_{it} = 3\right)\right).
\end{aligned}
$$

I assume that the nonpecuniary utility of enrolling in school

$$
\begin{aligned}
\kappa_S\left(hc_{it}, sub_{it}, s_{it}\right) = {} & \kappa_1 + \kappa_2 G_{it} + \kappa_3 \mathbb{1}\left(enroll_{i,t-1} = 1\right) + \kappa_4 \mathbb{1}\left(g_{i,t-1} = 1\right) \\
& +\kappa_5 \mathbb{1}\left(h_{it} = 1\right) + \kappa_6 \mathbb{1}\left(h_{it} = 2\right) \\
& +\kappa_7 \mathbb{1}\left(cigs_{it} = 2\right) + \kappa_8 \mathbb{1}\left(cigs_{it} = 3\right) \\
& +\kappa_9 \mathbb{1}\left(alc_{it} = 2\right) + \kappa_{10} \mathbb{1}\left(alc_{it} = 3\right) \\
& +\kappa_{11} \mathbb{1}\left(mj_{it} = 2\right) + \kappa_{12} \mathbb{1}\left(mj_{it} = 3\right).
\end{aligned}
$$

Lastly, I assume that the utility gained from using substances

$$\begin{aligned}
\alpha\left(sub_{it}, s_{it}\right) \quad = \quad & \left(\alpha_1 + \alpha_2 \mathbb{1}\left(\mathscr{C}_{it} = 0\right) + \alpha_3 \mathscr{C}_{it} + \alpha_4 \mathscr{A}_{it} + \alpha_5 \mathscr{M}_{it}\right) \mathbb{1}\left(cigs_{it} = 2\right) \\
& + \left(\alpha_6 + \alpha_7 \mathbb{1}\left(\mathscr{C}_{it} = 0\right) + \alpha_8 \mathscr{C}_{it} + \alpha_9 \mathscr{A}_{it} + \alpha_{10} \mathscr{M}_{it}\right) \mathbb{1}\left(cigs_{it} = 3\right) \\
& + \left(\alpha_{11} + \alpha_{12} \mathbb{1}\left(\mathscr{A}_{it} = 0\right) + \alpha_{13} \mathscr{C}_{it} + \alpha_{14} \mathscr{A}_{it} + \alpha_{15} \mathscr{M}_{it}\right) \mathbb{1}\left(alc_{it} = 2\right) \\
& + \left(\alpha_{16} + \alpha_{17} \mathbb{1}\left(\mathscr{A}_{it} = 0\right) + \alpha_{18} \mathscr{C}_{it} + \alpha_{19} \mathscr{A}_{it} + \alpha_{20} \mathscr{M}_{it}\right) \mathbb{1}\left(alc_{it} = 3\right) \\
& + \left(\alpha_{21} + \alpha_{22} \mathbb{1}\left(\mathscr{M}_{it} = 0\right) + \alpha_{23} \mathscr{C}_{it} + \alpha_{24} \mathscr{A}_{it} + \alpha_{25} \mathscr{M}_{it}\right) \mathbb{1}\left(mj_{it} = 2\right) \\
& + \left(\alpha_{26} + \alpha_{27} \mathbb{1}\left(\mathscr{M}_{it} = 0\right) + \alpha_{28} \mathscr{C}_{it} + \alpha_{29} \mathscr{A}_{it} + \alpha_{30} \mathscr{M}_{it}\right) \mathbb{1}\left(mj_{it} = 3\right) \\
& + \alpha_{31} age_{it} \mathbb{1}\left(cigs_{it} \geq 2\right) + \alpha_{32} age_{it} \mathbb{1}\left(alc_{it} \geq 2\right) + \alpha_{33} age_{it} \mathbb{1}\left(mj_{it} \geq 2\right) \\
& + \alpha_{34} \mathbb{1}\left(cigs_{it} \geq 2\right) \mathbb{1}\left(alc_{it} \geq 2\right) \\
& + \alpha_{35} \mathbb{1}\left(cigs_{it} \geq 2\right) \mathbb{1}\left(mj_{it} \geq 2\right) \\
& + \alpha_{36} \mathbb{1}\left(alc_{it} \geq 2\right) \mathbb{1}\left(mj_{it} \geq 2\right) \\
& + \alpha_{37} \mathbb{1}\left(alc_{it} = 2\right) \mathbb{1}\left(age_{it} \geq 21\right) + \alpha_{38} \mathbb{1}\left(alc_{it} = 23\right) \mathbb{1}\left(age_{it} \geq 21\right)
\end{aligned}$$

C Posterior Distribution

The posterior distribution is made up of the likelihood function and the prior distribution. For each individual i, I observe in the data a vector of optimal substance use and human capital accumulation choices $\left\{sub_{it}^{obs}, hc_{it}^{obs}\right\}_{t=1}^{T_i}$, a wage when an individual works, arrests, and grade completion status if the individual is enrolled in school. The additional latent variables that I simulate in order to augment the data are time persistent unobserved heterogeneity in preferences and skills μ_i; worker-firm match values $\psi_i\left(\cdot\right) = \left\{\psi_{ik}\right\}_{k=1}^{K_i}$; unobserved wage draws $w_i\left(\cdot\right) = \left\{\left\{w_{it}\left(hc_{it}\right)\right\}_{hc_{it} \neq hc_{it}^{obs}}\right\}_{t=1}^{T_i}$; unobserved parental transfers $\tilde{w}_i\left(\cdot\right) = \left\{\left\{\tilde{w}_{it}\left(hc_{it}\right)\right\}_{d_{it} \in D_{it}\left(s_{it}\right)}\right\}_{t=1}^{T_i}$, the latent variables that describe the probabilities of grade completion $G_i^*\left(\cdot\right) = \left\{G_{it}^*\right\}_{t=1}^{T_i}$ and arrest $R_i^*\left(\cdot\right) = \left\{R_{it}^*\right\}_{t=1}^{T_i}$; and the shocks to the substance use choice $\epsilon_i^{sub}\left(\cdot\right) = \left\{\epsilon_{it}^{sub}\right\}_{t=1}^{T_i}$. I denote the parameters to be estimated as θ. The likelihood function represents the probability of observing the augmented data conditional on the parameters. I assume diffuse priors on all of the parameters. I denote the density of μ_i, ψ_i, ϵ_{it}^{sub}, and $\tilde{w}_{it}\left(hc_{it}\right)$ as $g\left(\cdot|\theta\right),$ where $g\left(\cdot|\theta\right)$ is the normal pdf given the distributional parameters θ_1. Given this information, the joint posterior distribution of the parameters and the latent variables is proportional to

$$\Pr\left(\theta;\left\{\mu_i,\psi_i\left(\cdot\right),w_i\left(\cdot\right),\tilde{w}_i\left(\cdot\right),G_i^*,R_i^*,\epsilon_i^c\left(\cdot\right)\right\}_{i=1}^N\right)\propto \tag{19}$$

$$\prod_{i=1}^{N}\left[g\left(\mu_i|\theta\right)g\left(\psi_i|\theta\right)\right. \tag{20}$$

$$\cdot\prod_{t=1}^{T_i}\left[\mathbb{1}\left(R_{i,t}^*>0\;if\;r_{it}=1,<0\;if\;r_{it}=0\right)\cdot\Pr\left(R_{i,t}^*|s_{it},sub_{it}^{obs},\mu_i^{arrest};\eta_{arrest}\right)\right. \tag{21}$$

$$\cdot\left(\mathbb{1}\left(G_{i,t}^*>0\;if\;g_{it}=1,<0\;if\;g_{it}=0\right)\cdot\Pr\left(G_{i,t}^*|s_{it},sub_{it}^{obs},\mu_i^{pass};\eta_{school}\right)\right)^{\mathbb{1}\left(enroll_{it}=1\right)} \tag{22}$$

$$\cdot\,\mathbb{1}\left(\epsilon_{it}^{sub}\;is\;in\;bounds\right)\cdot g\left(\epsilon_{it}^{sub}|\theta\right) \tag{23}$$

$$\cdot\Pr\left(v_{it}\left(sub_{it}^{obs},hc_{it}^{obs}|s_{it},\mu_i,w_{it},\tilde{w}_{it},\epsilon_{it}^{sub};\theta\right)\geq v_{it}\left(sub_{it}\left(j\right),j|s_{it},\mu_i,w_{it},\tilde{w}_{it},\epsilon_{it}^{sub};\theta\right)\;\forall\,j\in D_{it}\left(s_{it}\right)\right) \tag{24}$$

$$\cdot\prod_{j=1}^{D_{it}(s_{it})}\left[\mathbb{1}\left(\tilde{w}_{it}\left(j\right)\;satisfies\;CONS_{it}\geq C_{min}\right)\cdot g\left(\tilde{w}_{it}\left(j\right)|\theta\right)\right. \tag{25}$$

$$\cdot\;\Pr\left(\ln\left(w_{it}\left(j\right)\right)|s_{it},\mu_i,\psi_i,sub_{it}\left(j\right);\theta\right)]]] \tag{26}$$

In the above equation, $\mathbb{1}\left(\cdot\right)$ is an indicator that equals one if the argument is true. The purpose of these terms is to make sure that the latent variables I simulate agree with the data. Line 21 guarantees that $R_{i,t+1}^*$ is greater than zero if an individual got arrested and is less than zero if he did not. Similarly the indicator in line 22 guarantees that $G_{i,t+1}^*$ is greater than zero if an individual completes a grade and less than zero if he did. Line 23 guarantees that ϵ_{it}^{sub} results in the observed substance use choice being the optimal continuous choice at the observed human capital accumulation choice. Lastly, line 25 guarantees that transfers are large enough so that consumption is larger than the minimum consumption level. As for the other terms in equation 19,

$$\Pr\left(R_{i,t+1}^*|s_{it},sub_{it}^{obs},\mu_i^{arrest};\eta_{arrest}\right)\;=\;\phi\left(R_{i,t+1}^*-\eta_{arrest}\left(\mu_i^{arrest},sub_{it}^{obs},s_{it}\right)\right)$$

and

$$\Pr\left(G_{i,t+1}^*|s_{it},sub_{it}^{obs},\mu_i^{pass};\eta_{school}\right)\;=\;\phi\left(G_{i,t+1}^*-\eta_{school}\left(\mu_i^{arrest},sub_{it}^{obs},s_{it}\right)\right),$$

where ϕ is the standard normal pdf.

Let

$$\tilde{v}_{it}\left(sub_{it}\left(j\right),j|s_{it},\mu_i,w_{it},\tilde{w}_{it},\epsilon_{it}^{sub};\theta\right)=v_{it}\left(sub_{it}\left(j\right),j|s_{it},\mu_i,w_{it},\tilde{w}_{it},\epsilon_{it}^{sub};\theta\right)-\epsilon_{it}^{hc}$$

Given this,

$$\Pr\left(v_{it}\left(sub_{it}^{obs}, hc_{it}^{obs}|s_{it}, \mu_i, w_{it}, \tilde{w}_{it}, \epsilon_{it}^{sub}; \theta\right) \geq v_{it}\left(sub_{it}\left(j\right), j|s_{it}, \mu_i, w_{it}, \tilde{w}_{it}, \epsilon_{it}^{sub}; \theta\right) \ \forall\, j \in D_{it}\left(s_{it}\right)\right)$$
$$= \frac{\exp\left(v_{it}\left(sub_{it}^{obs}, hc_{it}^{obs}|s_{it}, \mu_i, w_{it}, \tilde{w}_{it}, \epsilon_{it}^{sub}; \theta\right)\right)}{\sum_{j\in D_{it}(s_{it})} \exp v_{it}\left(sub_{it}^{obs}, hc_{it}^{obs}|s_{it}, \mu_i, w_{it}, \tilde{w}_{it}, \epsilon_{it}^{sub}; \theta\right)}.$$

Lastly, remember that today's wage offer is a function of today's substance use, which varies with the discrete choice. Therefore,

$$\ln\left(w_{it}\left(j\right)|s_{it}, \mu_i, \psi_i, sub_{it}\left(j\right); \theta\right) \sim N\left(\theta_w^h\left(sub_{it}\left(j\right), s_{it}\right) + \psi_{ik} + \mu_i^w., \sigma_h^2\right).$$

So,

$$\Pr\left(\ln\left(w_{it}\left(j\right)\right)|s_{it}, \mu_i, \psi_i, sub_{it}\left(j\right); \theta_w\right) \propto$$
$$\sigma_h^{-1} \exp\left(-\frac{1}{2}\sigma_h\left(w_{it}\left(j\right) - \mu_{it}^h\left(j\right)\right)^2\right).$$

I divide the parameters and latent variables into blocks and draw from each block using the Metropolis-Hastings algorithm.