Persistent Marijuana Use: Evidence from the NLSY

Jorgen Hansen* and Golnaz Davalloo†

Abstract

We analyze persistence in marijuana consumption utilizing data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97). We allow for three sources of persistence: pure state dependence, time invariant unobserved heterogeneity and persistence in idiosyncratic, time-varying shocks. We also consider intensity of consumption based on days of use per month and estimate a dynamic ordered Probit model using simulated Maximum Likelihood. We consider a Polya model that generalizes the more commonly used Markov models. The results show that there is a causal effect of previous use. However, ignoring unobserved heterogeneity and serially correlated shocks significantly exaggerates the state dependence.

JEL Code: I12,I21

Keywords: marijuana; persistence; state dependence; unobserved heterogeneity; dynamic ordered probit; simulation; NLSY

*Concordia University, CIRANO, CIREQ and IZA. Corresponding author. Email: Jorgen.Hansen@concordia.ca

†Concordia University. Email: golnaz.davalloo@gmail.com

1 Introduction

The legal status of recreational marijuana in the US has changed significantly since 2012 when Colorado and Washington became the first states to legalize cannabis for adult use. Currently, recreational use is legal in as many as 18 states plus the District of Columbia. These changes have occurred despite evidence pointing to negative impacts from marijuana use (especially at young ages) on different outcomes, such as educational attainment (Davalloo and Hansen, 2022), school to work transitions (Williams and van Ours, 2020), financial and relational difficulties in adulthood (Chan et al., 2021; Cerda et al, 2016), health (Hall and Degenhardt, 2009; Lev-Ran et al., 2014), and welfare use and unemployment (Fergusson and Boden, 2008; Schmidt et al., 1998). Marijuana consumption has also been shown to increase the risk of consuming hard drags (see Deza, 2015).

It is however possible that the nature of marijuana consumption, and its associated risks, is heterogeneous in the population. For many, consumption is modest, occasional and highly transitory while others use marijuana on a regular and persistent basis, and the existence and magnitude of any negative impacts of marijuana use is likely to vary with consumption patterns. However, if there is a causal, addictive effect of marijuana use over time, any initiation is associated with a risk of continued, persistent use. In this case, policies that make marijuana consumption more accessible and socially acceptable may therefore increase the risk of marijuana dependence. On the other hand, if there is no causal effect of past marijuana use on current consumption, this risk is eliminated. It is therefore important to understand the dynamics of marijuana consumption and how it varies, at an individual level, over time.

In this paper we analyze transitions into and out-of marijuana consumption. Data from the 1997 cohort of National Longitudinal Survey of Youth (NLSY97) show that the probability of using marijuana in a given year is almost two times higher for those who used it the year before compared to those who did not use it. However, this data pattern is uninformative about the nature of marijuana persistence. Does past consumption cause current use (perhaps by changing preferences for the drug)? Or is the data simply reflecting different innate propensities to use marijuana over time where some youth receive substantial utility from marijuana consumption and therefore continuously use it while others receive a negative utility and never use it. A third possibility for the observed time dependence is persistence in random shocks to the utility of consumption. For example, an event in school or within the family may alter the perceived the utility and induce consumption in a given year. This effect may then persist over time. Our aim in this paper is to estimate the sources for persistence in marijuana consumption and evaluate their relative importance for overall persistence.

Our empirical framework builds on the influential work by Heckman (1981) and others who have developed models designed to separate true state dependence from spurious dependence (due to persistent unobserved heterogeneity). These models have been estimated for a number of different outcomes, such as welfare (Card and Hyslop, 2005; Hansen and Lofstrom, 2009), labor supply (Hyslop, 1999), unemployment (Hansen and Lofstrom, 2009) and health (Carro and Traferri 2014). A particularly relevant study for this paper is Deza (2015) who use a dynamic discrete choice model to analyze persistence in illicit drug use. Using data from the 1997 cohort of the NLSY, she estimates a general model of alcohol, marijuana and hard drug use and separate the contributions from state dependence and unobserved heterogeneity, both within drugs but also between drugs. Her results show the existence of significant "stepping-stone" effects into hard drugs, where current alcohol and marijuana use significantly increase the probability of hard drug use in the future.

Our paper addresses some important shortcomings in the previous literature. We first analyze the probability of marijuana use among American youth from ages 13 to

26, paying particular attention to its persistence. Apart from Deza (2015), there are few studies that have analyzed time dependence or persistence in marijuana consumption. While Deza (2015) estimates a general, dynamic model of consumption of alcohol and hard drugs, in addition to marijuana, the focus is on structural state dependence and transitions from alcohol and marijuana into hard drugs (that is, if softer drugs serve as "stepping-stones" into hard drugs). Our model specification, while limited to marijuana consumption only, allows for more general forms of dynamics as well as serially correlated utility shocks. We also estimate different persistence probabilities conditional on the amount consumed, allowing for the separation of occasional or experimental use from continuous, intensive use. We show that these additional dimensions are important and that moderate consumption of marijuana may serve as a "stepping-stone" into heavy use.[1]

The results indicate that serial correlation in the time-varying utility shocks contributes substantially to overall, observed persistence. If ignored, the estimate for structural state dependence and the estimated variance of persistent unobserved heterogeneity are exaggerated, leading to incorrect inference about sources of persistence. Further, separating moderate use from intense use is important.

Focusing first on the estimated average partial effects, which are designed to show the causal effect of past consumption on current consumption, our results for the most general specification of the binary case suggest that consumption of marijuana in the previous period increases the probability of current consumption by 0.129.[2] Given an unconditional consumption rate of 15-20 percent (depending on age), this effect is very

[1] We define moderate use as consumption less than 9 times per month and heavy use as 10 days or more of consumption. The data show that persistence is concentrated among heavy users while moderate use is more transitory. Specifically, the average probability of heavy marijuana use is 0.164, conditional on moderate consumption in the previous time period. This should be compared to a probability of 0.021 among those who did not use marijuana in the previous period.

[2] The average partial effect is estimated as $\hat{Pr}\left(y_{i,t}=1|y_{i,t-1}=1\right)-\hat{Pr}\left(y_{i,t}=1|y_{i,t-1}=0\right)$, which is averaged across individuals and time periods.

large. However, it is still significantly smaller than the corresponding effect obtained from a one-period lagged Markov model (where the effect is 0.192).

For the ordered model, we estimate two average partial effects for each intensity level. For moderate consumption levels, the first effect is the difference in conditional probabilities of moderate consumption when we condition on moderate versus no consumption in the previous time period while the second effect conditions on moderate and heavy use instead. The former effect (moderate versus no consumption) is 0.046 while the second effect is -0.051. That is, the probability of consuming moderate levels of marijuana in year t is 4.6 percentage points higher if the person consumed the same level of marijuana in year t-1, relative to not using any marijuana in year t-1. While the magnitude of this effect is smaller than the one obtained in the binary case, it constitutes a relative effect that is close to 50 percent, given the observed proportions of moderate consumption that are observed in the data. The negative effect for moderate versus heavy usage suggests a higher probability of moderate use in year t for those with a heavy consumption in the previous year compared to those with moderate consumption.

For heavy consumption levels, the first effect is the difference in conditional probabilities of heavy consumption when we condition on heavy versus no consumption in the previous time period while the second effect conditions on heavy and moderate use instead. The former effect equals 0.043 and is similar to the one estimated for moderate use. The second effect is smaller, 0.027. That is, the probability of consuming heavy levels of marijuana in year t is 4.3 percentage points higher if the person consumed the same level of marijuana in year t-1, relative to not using any marijuana in year t-1. Again, while the magnitude of this effect is smaller than the one obtained in the binary case, it constitutes a relative effect that is close to 50 percent, given the observed moderate consumption rates observed in the data.

Finally, our analysis of the sources for persistence in marijuana consumption reveals some interesting patterns. In the binary case, 52 percent of the persistence is causal (true state dependence). The remaining sources for the time dependence in marijuana consumption are: i) persistence in time-varying utility shocks (18 percent); ii) persistent, observed individual characteristics, such as race, gender and family background (16 percent); and lastly iii) persistent, unobserved heterogeneity (14 percent).

The estimated persistence probabilities for the ordered model suggest that time-invariant, unobserved heterogeneity plays a larger role for persistence of intense marijuana consumption (40 percent of overall persistence is due to unobserved heterogeneity) and less so for moderate use (32 percent). Persistence in time-varying utility shocks and persistence due to time-invariant observed individual characteristics play a similar role to that obtained in the binary mode. Moreover, for moderate use, true or causal state dependence accounts for 47 percent of total persistence while it is less important for heavy consumption levels (33 percent). That is, most of the overall persistence in moderate consumption is due to structural state dependence (this result also applies when we consider consumption as a binary outcome) while for heavy consumption, most of the persistence is due to individual heterogeneity.

The rest of the paper is organized as follows. In the next section, we describe the data and in Section 3 we present the econometric model and its results when we consider marijuana consumption as a binary outcome. Section 4 is structured similarly but for the generalized model with ordered outcomes. Section 5 concludes the paper with a brief summary.

2 Data

In this paper, we utilize data from the 1997 cohort of the National Longitudinal

Survey of Youth (NLSY97), which is a nationally representative sample of five cohorts of males and females who were born between 1980 and 1984. The initial interview took place in 1997 and follow-up interviews were conducted annually until 2011 after which it became a biannual survey. NLSY97 gathers information in an event history format, in which dates are collected for the beginning and end of significant life events. In addition, there are detailed information on family background and income as well as on individual scholastic ability.

In our analysis, we remove individuals who were not part of the representative cross-sectional sample in 1997 (this removes oversamples of Blacks and Hispanics). In order to reduce potential initial conditions concerns, we also exclude all respondents who were born before 1983. Most of those born in 1983 were 13 years old at the time of the first survey while most of those born in 1984 were 12 years old at that interview. We are then left with 1,589 individuals. Of these, 55 reported having used Marijuana before the age of 13 and to avoid left censoring, these were removed.

We also excluded individuals who did not provide valid information on the following: family income (at any point between 1997 and 2001), mother's age at birth, family situation at the time of the survey (divorced parents or not), area of residence, number of siblings, mother's education and Armed Forces Qualification Test (AFQT) scores.[3] We exclude those with missing information on any of these variables since they are included as covariates in all model specifications.[4] Finally, we remove respondents who did not provide any answers on questions related to marijuana use and those who we

[3]AFQT scores consists of four components of the Armed Forces Vocational Aptitude Battery (ASVAB): Arithmetic Reasoning (AR), Mathematics Knowledge (MK), Word Knowledge (WK), and Paragraph Comprehension (PC). These scores have been used extensively in research on education using NLSY data. In this paper, we follow Belzil and Hansen (2020) and regress the scores on age and education, in order to adjust for age and educational differences at the time of the test, and use the standardized residual from that regression as the measure of cognitive ability.

[4]These variables are commonly included in empirical analysis of substance use. We decided not to include father's education in the list mainly because of the large number of missing values for this variable and the skewness in responses to questions about this across the sample (there is a higher fraction of missing among non-white respondents).

only observed once. After these selections, the sample consists of 1,204 individuals.

We use information on family income for each individual at ages 16 and 17, if available, and construct an average income measure. If income is only available for one of the years, the average income is replaced by that income. If no income information is available for these ages, we consider income at earlier ages if available in order to minimize the number of individuals dropped because of missing income. We express income in year 2000 dollars using the CPI for all urban consumers.

To derive measures of marijuana use, we compile information from questions like: 1) have you ever used marijuana?; 2) when did you start using marijuana?; 3) did you use marijuana during the year before the interview? and 4) On how many days have you used marijuana in the last 30 days? From the responses to these questions, we create individual annual indicators of marijuana use (and non-use) as well as indicators for intensity of use, conditional on use (less than 10 days last month versus 10 days or more). Responses to the first three questions are used to validate consistency in responses while our outcome variables are derived from answers to the fourth question.

In Table 1 below, we present the proportions of the sample that used marijuana at a given age. At age 13, 3.7 percent of the respondents used marijuana at least once. Three years later, at age 16, this had increased almost fivefold to 18.3 percent. After 16, the proportion of users increase until age 18 when it peaks and then declines to around 16 percent when respondents are in their 20s.

The entries in Table 1 do not reveal how respondents move in and out of marijuana use. In order to infer the degree of time persistence and the transitory nature of marijuana use, we show average (across individuals and time periods) conditional probabilities in Table 2. The entries show row percentages of the probability of using marijuana in year t, conditional on marijuana use in year t-1. The top row entries show that 91.5 percent of those who did not use marijuana in year t-1 continued to be non-users in year t, while 8.5 percent started using marijuana. Similarly, among those who used marijuana in year t-1, 63 percent continued using it in year t while 37 percent stopped.[5]

While the entries in Table 1 show how usage vary with age, the entries in Table 2 show the anatomy of usage in any year. That is, how many start using it and how many stop. The focus of this paper is to analyze the persistence over time in marijuana use and estimate to what extent it is causal (or due to addiction) as opposed to persistence in observed and unobserved characteristics.

In Table 3, we show average characteristics separately for individuals who never used marijuana and for those who used it at least once over the sample period. Overall, males and Hispanics are somewhat overrepresented among users. The proportion living with both biological parents at the interview date is higher among the never-users (0.66) than among the users (0.57). For other background variables - family income, mother's education, AFQT scores, mother's age at birth, urban residence and number of siblings - there are no major differences in sample means between the two groups. Lastly, half of our sample have used marijuana at least once. This is somewhat lower than the 57-58 percent reported in Deza (2015).

Similar to earlier studies on substance use that utilize retrospective information, our measures of marijuana are subject to potential measurement error problems, specifically recall errors. However, unlike most of them (see for instance Van Ours and Williams (2009) whose sample consists of respondents aged 25-50), the respondents in our sample were first asked about their marijuana use at a young age (age 12 or 13). We therefore believe the issue of recall errors is less serious in this paper than in many of the previous studies on this topic.

[5] Deza (2015) reports similar proportions (an entry probability of 9.2 percent and a persistence probability of 67 percent (Table 2, panel B)) using NLSY97, despite different sample selections. She limited her sample to respondents with a valid state of residence at each wave between 1997 and 2007, i.e. a balanced panel. She also included the oversample of minorities available in NLSY97.

3 Binary outcomes

3.1 Estimation

In this paper we explore the persistence in marijuana use and its sources. Exploiting the longitudinal nature of the NLSY97 data, we analyze the dynamics of marijuana use (and non-use). Our empirical models are inspired by Heckman (1981) who derived a general framework for the analysis of discrete choices in discrete time. He showed that observed choices can be derived from latent variables, which in turn can be thought of as describing utility differences across alternatives. Hence, observed choices are outcomes of utility maximization. We follow Lee (1997) and Liu et al (2012) who offers a description and assessment of generalized versions of Heckman's original framework.

Specifically, let y^*_{it} denote latent, unobserved utility differences, for individual i in period t, between using and not using marijuana

$$y^*_{i,t} = \Psi_{i,t} + \gamma y_{i,t-1} + \sigma\mu_i + \varepsilon_{i,t} \tag{1}$$

for $i = 1, ..., n; t = 1, ..., T_i$ and where $\Psi_{i,t} = X_i\beta + \kappa_1 (t - t_0) + \kappa_2 (t - t_0)^2$. If the utility difference is positive, individual i consumes marijuana in period t and the observed outcome is

$$y_{i,t} = \begin{cases} 1 \; if \; y^*_{i,t} > 0 \\ 0 \; if \; y^*_{i,t} \leq 0 \end{cases}$$

In our case, $y_{i,0} = 0$ as we start observing and modeling marijuana use at age 13. We include a fairly rich set of observable characteristics in X and assume that the error terms (μ_i) and ($\varepsilon_{i,t}$) are independent of X and across individuals. While μ_i is fixed over time, $\varepsilon_{i,t}$ is time-varying and possibly correlated over time. There are four possible sources of time persistence in marijuana use in equation (1): i) time-invariant observed characteristics (X_i); ii) true state dependence ($\gamma > 0$); iii) time-invariant unobserved characteristics (μ_i); and iv) persistence in time-varying shocks ($\varepsilon_{i,t}$).

In equation (1), it is assumed that the dynamics of marijuana use can be fully captured by lagged choices ($y_{i,t-1}$). Alternatively, we can imagine that there is some memory in the process and that usage in previous periods may also have a direct or causal impact on current use. To allow for this, we consider a more general dynamic representation, described as the Polya model in Lee (1997), where the latent variable $y^*_{i,t}$ is expressed as:

$$y^*_{i,t} = \Psi_{i,t} + \gamma \sum_{j=1}^{t} \delta^{j-1} y_{i,t-j} + \sigma\mu_i + \varepsilon_{i,t} \tag{2}$$

for $i = 1, ..., n; t = 1, ..., T_i$ and where $\delta,\ [0, 1]$ can be thought of as a discount factor. When $\delta = 0$, past choices beyond t-1 do not matter for the utility in period t whereas when $\delta = 1$, the impact of past choices do not fade with time.

We assume that $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + \nu_{i,t}$, where ν_{it} are i.i.d $N(0, 1)$, and consequently the choice probabilities involve multiple integrals. Following Lee (1997), we adopt the Geweke-Hajivassiliou-Keane (GHK) simulator and estimate the parameters in equations (1) and (2) using Maximum Simulated Likelihood. The joint probability for observed choices $y_{i,1}, ..., y_{i,T}$, conditional on X_i and μ_i is

$$Pr\left(y_{i,1}, ., y_{i,T}|X_i, \mu_i\right) = \int_{L_1}^{U_1} . \int_{L_T}^{U_T} f\left(\varepsilon_{i,T}|\varepsilon_{i,T-1}, ., \varepsilon_{i,1}\right) f\left(\varepsilon_{i,T-1}|\varepsilon_{i,T-2}, ., \varepsilon_{i,1}\right) ... f\left(\varepsilon_{i,1}\right) d\varepsilon_T . d\varepsilon_1 \tag{3}$$

where $f\left(\varepsilon_{i,t}|\varepsilon_{i,t-1}, .., \varepsilon_{i,1}\right)$ is the density of $\varepsilon_{i,t}$ conditional on past realizations of ε and the integral limits are

$$L_t = \begin{cases} -\left(\Psi_{i,t} + \gamma \sum_{j=1}^{t} \delta^{j-1} y_{i,t-j} + \sigma\mu_i\right) & if\ y_{i,t} = 1 \\ -\infty & if\ y_{i,t} = 0 \end{cases}$$

and

$$U_t = \begin{cases} \infty & if\ y_{i,t} = 1 \\ -\left(\Psi_{i,t} + \gamma \sum_{j=1}^{t} \delta^{j-1} y_{i,t-j} + \sigma\mu_i\right) & if\ y_{i,t} = 0 \end{cases}$$

Lee (1997) shows how the joint probability in (3) can be expressed using standard normal density and distribution functions and simulated using the GHK simulator. The sample likelihood then becomes

$$\mathcal{L} = \sum_{i=1}^{n} ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T_i} \Phi \left(D_{i,t} \left(\Psi_{i,t} + \gamma \sum_{j=1}^{t} \delta^{j-1} y_{i,t-j} + \sigma\mu_i^j + \rho\varepsilon_{i,t-1}^j \right) \right) \right\} \quad (4)$$

where $D_{i,t} = 2y_{i,t} - 1$. The random disturbances $\varepsilon_{i,t}$ are recursively generated as described in Lee (1997).[6] The $\mu's$ are generated from $N(0,1)$ random draws while the $\varepsilon's$ are generated from functions of $U[0,1]$ draws. Lee (1997) provides Monte Carlo results for this and other dynamic specifications and concludes that this estimator generally performs well. Since we use an unbalanced panel, T_i varies between 2 and 14. We set $m = 100$.

[6]We provide a description of the generation of truncated random draws needed for the likelihood function in the Appendix.

3.2 Empirical results

In this section, we present both parameter estimates and average partial effects of selected variables. We use a parametric bootstrap to estimate the standard errors of the average partial effects. Specifically, for each model we draw 100 vectors of parameter values from the estimated variance-covariance matrix. For each vector and variable of interest, we calculate a partial effect. The reported effects below are the average effects across the 100 draws and the standard errors of the effects are estimated using the standard deviation of the simulated effects.

3.2.1 Estimates and average partial effects

Estimates from three alternative Probit specifications are presented in Table 4. This will allow us to analyze how the average partial effects depend on stochastic assumptions and specifications of the dynamic relationship of marijuana consumption.

The entries in column one refer to a specification where dynamics in marijuana use is represented by a first-order Markov but with no time-invariant unobserved heterogeneity and no persistence in the time-varying shocks. In column two, we retain the assumption of a first-order Markov but allow for both unobserved heterogeneity and serial correlation in the time-varying shocks. Finally, in column three we generalize dynamics of marijuana use by incorporating marijuana use from periods before last year (see equation 2 above). We set δ to 0.7.

There is evidence of significant time dependence in marijuana use. The estimate in column one for marijuana use in the previous period (γ) is 1.691 and it is statistically significant. However, as discussed above, in this simplified model, all persistence in marijuana is captured by this parameter and it is therefore unlikely to represent the true (or causal) effect of past use on current use. Maintaining the same dynamic structure but allowing for another source of persistence has a dramatic (and expected) effect.

The estimate in column two is 0.976, suggesting that the causal effect of past usage is seriously exaggerated in the naive specification in column one. Instead, a significant part of the observed persistence is due to time-invariant, unobserved heterogeneity with $\hat{\sigma}$ equal to 0.851.

The corresponding estimates reported in column three suggest important roles for all three sources of time dependence. The estimate of previous use (γ) is further reduced to 0.732 while $\hat{\sigma}$ equals 0.414. Further, $\hat{\rho}$ is significant and equals 0.220. At the bottom of Table 4, we report the Akaike Information Criteria (AIC) for each model specification and these favor the most general model presented in column three.

Regarding observable characteristics, the entries in Table 4 suggest that gender, family stability and size, cognitive skills and peer effects matter for marijuana use. The estimates associated with these variables are significant and generally similar across all three specifications while the estimates of the other included variables (shown in Table A1) are not.

In Table 5 we show the average partial effects for selected variables. The average partial effects are estimated as $\hat{Pr}\left(y_{i,t}=1|y_{i,t-1}=1\right)-\hat{Pr}\left(y_{i,t}=1|y_{i,t-1}=0\right)$, and they are averaged across individuals and time periods. The first row shows the predicted difference in the probability of using marijuana between users and non-users in the previous period. According to these estimated effects - for the restrictive model with a first-order Markov dynamics, no unobserved heterogeneity and no serial persistence in the error terms - the probability of marijuana use in any given year is 47 percentage points higher if the person used marijuana the year before. This is a very large effect considering that the proportion of the sample that use marijuana at any given age very between 15 and 20 percent (after age 14, see Table 1). However, as we generalize the models, this conditional probability is reduced. In column two, the difference is 19.2

percentage points while in column three it has been reduced to 12.9 percentage points.[7]

The remaining entries in Table 5 show estimated marginal effects of the variables whose parameter estimates are statistically significant. Overall, and unlike the effect of past use, the magnitudes are similar across the different model specifications. For instance, the predicted probability of using marijuana is around two percentage points higher for males than for females while it is around two percentage points lower for students living with both biological parents at the time of the interview. Students with higher cognitive test scores (AFQT) have higher predicted probabilities of marijuana use although the differences are small (a one standard deviation increase in test scores raise the probability with less than one percentage point). Finally, the effect of peers is just over one percentage point across all specifications suggesting that students with favorable peers are less likely to use marijuana.

3.2.2 Model fit

We assess the model's ability to generate outcomes that match those observed in the data by predicting transition probabilities. In Table 6, we show the predicted transition matrix for marijuana use obtained by simulating outcomes generated by the estimates from the general Polya model (Model 3 in Table 4). The predicted conditional probabilities, which are averaged over individuals and time, match those in the data (presented in table 2) well. For example, the probability of using marijuana in year t, conditional on using marijuana in year t-1, is 0.63 in the data and the predicted probability is 0.66. Moreover, the probability of using marijuana in year t, conditional on *not* using marijuana in year t-1 is 0.085 in the data while the predicted probability is 0.099.

[7] The average partial effect for Model 2 is a bit lower than the corresponding effect (25.1 percentage points) reported in Deza (2015). Her model, like the one in Model 2, ignores serial persistence in utility shocks and assume that a first-order Markov structure accurately captures dynamics in marijuana consumption.

3.2.3 Sources of persistence

In Table 7 we explore the anatomy of persistent marijuana use. The entries are obtained using estimates from the Polya model and in the first row, we replicate the the probability of using marijuana in year t, conditional on using marijuana in year t-1, from Table 6. This is the predicted persistence. In the second row, we remove the role of time-invariant unobserved heterogeneity by setting $\sigma = 0$ and the predicted probability drops from 0.661 to 0.567. Thus, removing time-invariant unobserved heterogeneity reduce persistence with 14 percent. In row three, we remove persistence in the time-varying utility shocks by setting $\rho = 0$ (in addition to setting $\sigma = 0$). The predicted persistence further drops to 0.449 indicating that this source of persistence contributes about 20 percent to the overall persistence.

Finally, in the last row, we also remove the effect of time-invariant observed characteristics and the time trend by setting $\beta = \kappa_1 = \kappa_2 = 0$ (in addition to fixing $\sigma = \rho = 0$). This further reduce the persistence from 0.449 to 0.345. The remaining persistence (52 percent of the total) is due to a causal or addictive effect of using marijuana in the previous period. Thus, a majority of the observed state dependence in marijuana consumption is causal although a large portion is due to persistence in utility shocks and heterogeneity. A similar finding is reported in Deza (2015).

4 Ordered outcomes

The results so far are based on the dichotomy of marijuana use with no separation between occasional or moderate consumption and more intense, regular use. This is arguably restrictive and to allow for different effects depending on the intensity of consumption, we generalize the model described above to include multiple, ordered outcomes.[8]

4.1 Estimation

Specifically, let $c_{i,t}^{*}$ denote latent, unobserved utility of marijuana consumption for individual i in period t

$$c_{i,t}^{*} = \Psi_{i,t} + \gamma_1 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 1\right) + \gamma_2 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 2\right) + \sigma\mu_i + \varepsilon_{i,t} \qquad (5)$$

for $i = 1, ..., n; t = 1, ..., T_i$ and where $\Psi_{i,t} = X_i\beta + \kappa_1 (t - t_0) + \kappa_2 (t - t_0)^2$. $1(.)$ is an indictor function that equals one if the argument is true and zero otherwise. If utility is below a certain level (θ_1), the individual is not consuming marijuana in period t. If utility exceeds (θ_1) but is below (θ_2), the individual consumes a moderate amount of marijuana in period t and finally, if utility exceeds (θ_2), the individual is a heavy user. Thus, the observed outcome $(c_{i,t})$ is

$$c_{i,t} = \begin{cases} 0 \; if \; c_{i,t}^{*} \leq \theta_1 \\ 1 \; if \; \theta_1 < c_{i,t}^{*} \leq \theta_2 \\ 2 \; if \; c_{i,t}^{*} > \theta_2 \end{cases}$$

As mentioned above in the binary case, $c_{i,0} = 0$ since we start observing and modeling marijuana use at age 13. We maintain the assumptions that the error terms (μ_i) and (ε_{it}) are independent of X and across individuals, μ_i is i.i.d. $N(0, 1)$ and fixed

[8] Honore et al (2021) derive a generalized method of moments estimator for a dynamic ordered Logit model with fixed effects, assuming time independence of the utility shocks. We argue that since we observe the initial conditions, the argument for using a fixed effects estimator instead of a random effects estimator (like we do) is weaker.

over time while $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ are i.i.d $N(0,1)$. We define $c_{i,t} = 0$ if the person did not use marijuana in period t, $c_{i,t} = 1$ if the person used marijuana less than 10 times per month in period t (moderate use) and $c_{i,t} = 2$ if the person used marijuana 10 times or more per month in period t (heavy use).

Given the stochastic assumptions and the assignment rule above, the probabilities of observed outcomes are then

$$\begin{aligned}
Pr\left(c_{i,t} = 0 | c_{i,t-1}\right) &= \Phi\left(\theta_1 - \lambda_{i,t}\right) &&= \Lambda_0 \\
Pr\left(c_{i,t} = 1 | c_{i,t-1}\right) &= \Phi\left(\theta_2 - \lambda_{i,t}\right) - \Phi\left(\theta_1 - \lambda_{i,t}\right) &&= \Lambda_1 \\
Pr\left(c_{i,t} = 2 | c_{i,t-1}\right) &= 1 - \Phi\left(\theta_2 - \lambda_{i,t}\right) &&= \Lambda_2
\end{aligned}$$

where

$$\lambda_{i,t} = \Psi_{i,t} + \gamma_1 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 1\right) + \gamma_2 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 2\right) + \sigma\mu_i + \rho\varepsilon_{i,t-1}$$

We again adopt the Geweke-Hajivassiliou-Keane (GHK) simulator and estimate the parameters in equation (5) using Maximum Simulated Likelihood. The sample likelihood is an adjusted version of the one presented in equation (4) above

$$\mathcal{L} = \sum_{i=1}^{n} \; ln\left\{\tfrac{1}{m}\textstyle\sum_{j=1}^{m}\prod_{t=1}^{T_i} {\Lambda_0}^{I(c_{it}=0)} \Lambda_1^{I(c_{it}=1)} \Lambda_2^{I(c_{it}=2)}\right\} \tag{6}$$

The random disturbances $\varepsilon_{i,t}$ are generated recursively, similar to the binary case, and the $\mu's$ are generated from $N(0,1)$ random draws while the $\varepsilon's$ are generated from functions of $U[0,1]$ draws.[9] We set $m = 100$.

4.2 Empirical results

4.2.1 Descriptive statistics

The proportions of the sample that used marijuana at a given age, by intensity level, are presented in Table 8. At age 13, of the 3.7 percent of the respondents who used marijuana at least once, a majority (73 percent) used it occasionally (less than 10 days during the 30 days preceding the survey date). Three years later, at age 16, the proportion of intense users, among all users, increase to 33 percent. In fact, the proportion of intense users, among all users, increase with age and reach over 60 percent at age 26. This suggests a higher degree of persistence among the intense users.

The entries in Table 9 show the degree of time persistence and the transitory nature of marijuana use, conditional on intensity of consumption. Like before, we show average (across individuals and time periods) conditional probabilities and the entries show row percentages of the probability of consuming a certain level of marijuana in year t, conditional on marijuana use in year t-1. The top row entries show, like before, that 91.5 percent of those who did not use marijuana in year t-1 continued to be non-users in year t. Among the remaining non-users, 6.4 percent started consuming marijuana at a moderate intensity level while 2.1 percent (a quarter of those who started using marijuana) used marijuana intensively (used it at least 10 days or more during the 30 days preceding the survey date). Among those who used marijuana moderately in year t-1, almost half stopped consuming it in year t while 16 percent increased their consumption the following year. Only 34 percent continued with moderate use, suggesting a transitory nature among occasional or moderate users. The entries in the last row show that 20 percent of the intense users in period t-1 stopped using marijuana in

[9] See the Appendix for details.

period t while 16.6 percent reduced their consumption (but kept consuming). However, the majority (63.5 percent) continued their intense level of consumption the following year (in year t).

4.2.2 Estimates and average partial effects

Estimates from the ordered Probit Polya model (the likelihood presented in equation 6) are shown in Table 10. Similar to the binary case, we set δ to 0.7. The model includes the same set of observed characteristics as the ones for the binary case but we report only a subset of the associated estimates in Table 10 (those that are statistically significant). The full set of estimates are provided in Table A2 in Appendix.

The estimates in the first two rows suggest existence of true or causal time dependence in outcomes and this dependence is stronger for intense marijuana use. The estimates are 0.432 and 0.786 for moderate and heavy use, respectively. We will illustrate how these estimates translate into average partial effects and predicted transition probabilities below. The estimates for male, intact family and peers are similar in magnitude (and statistical significance) to those obtained in the binary case (see column 3 of Table 4). The standard deviation of the persistent unobserved heterogeneity term, $\hat{\sigma}$, is 0.569, again similar to the estimate in the binary model. Finally, there is evidence of serial persistence in the error terms (ε_{it}) as $\hat{\rho}$ is significant and equals 0.300.

In Table 11 we show the average partial effects for selected variables. The first two rows show the predicted difference in the probability of using marijuana at a moderate level when we condition on different consumption levels in the previous time period. The first effect is the difference in conditional probabilities of moderate consumption when we condition on moderate versus no consumption in the previous time period while the second effect conditions on moderate and heavy use instead. The former effect (moderate versus no consumption) is 0.046 while the second effect is -0.051.

That is, the probability of consuming moderate levels of marijuana in year t is 4.6 percentage points higher if the person consumed the same level of marijuana in year t-1, relative to not using any marijuana in year t-1. While the magnitude of this effect is smaller than the one obtained in the binary case, it constitutes a relative effect that is close to 50 percent, given the observed moderate consumption rates observed in the data. The negative effect for moderate versus heavy usage suggests a higher probability of moderate use in year t for those with a heavy consumption in the previous year compared to those with moderate consumption.

In rows three and four we present the corresponding probability differences for heavy consumption levels. The first effect is the difference in conditional probabilities of heavy consumption when we condition on heavy versus no consumption in the previous time period while the second effect conditions on heavy and moderate use instead. The former effect equals 0.043 and is similar to the one estimated for moderate use. The second effect is smaller, 0.027. That is, the probability of consuming heavy levels of marijuana in year t is 4.3 percentage points higher if the person consumed the same level of marijuana in year t-1, relative to not using any marijuana in year t-1. Again, while the magnitude of this effect is smaller than the one obtained in the binary case, it constitutes a relative effect that is close to 50 percent, given the observed moderate consumption rates observed in the data.

The remaining entries in Table 11 show estimated marginal effects of the variables whose parameter estimates are statistically significant. For all four variables (male, intact family, afqt and peers), the average partial effects are larger in absolute value for moderate use than for heavy use. For example, the predicted probability of using a moderate level of marijuana is 1.3 percentage points higher for males than for females while it is only 0.4 percentage points higher in the heavy consumption case. A similar difference applies to the impact of living with both biological parents at the time of the interview. While youth in intact families are less likely to use any level of marijuana, the strength of the effect is weaker for heavy use (-0.005 versus -0.016 for moderate use).

4.2.3 Model fit

Similar to the binary case presented above, we assess the model's ability to generate outcomes that match those observed in the data by predicting transition probabilities. In Table 12, we show the predicted transition matrix for marijuana use obtained by simulating outcomes generated by the estimates from the ordered Polya model. The predicted conditional probabilities, which are averaged over individuals and time, match those in the data (presented in Table 9) reasonably well. For example, the probability of not using marijuana in year t, conditional on not using marijuana in year t-1 is 0.915 in the data and the predicted probability is 0.92. The predicted entry probabilities, going from non-use to moderate or intense use, also match those in the data well.

The second row entries show probabilities of various use conditional on moderate use in period t-1. The predicted exit (or stopping) probability is 0.551 compared to 0.497 in the data. However, the model underestimates the probability of remaining a moderate user somewhat (0.254 versus 0.339) and slightly exaggerates the transition from moderate to intense use (0.195 versus 0.164). Conditional on heavy use, the predicted probabilities are similar to those in the data, especially the probability of remaining an intense user (0.598 versus 0.635 in the data). Overall, the model generates predicted transition matrix entries that match those in the data well.

4.2.4 Sources of persistence

In Table 13 we replicate the analysis on the anatomy of persistent marijuana use but generalize it to allow differential impacts on moderate and heavy use. The entries in column one refers to moderate use, $\hat{Pr}\left(y^m_{i,t} = 1 | y^m_{i,t-1} = 1\right)$, while those in column two refer to heavy use, $\hat{Pr}\left(y^h_{i,t} = 1 | y^h_{i,t-1} = 1\right)$. They are obtained using estimates from the ordered Polya model and in the first row, we replicate the the probabilities of marijuana consumption in year t, conditional on the same intensity level of marijuana consumption in year t-1, from Table 12. In the second row, we remove the role of time-invariant unobserved heterogeneity by setting $\sigma = 0$. The predicted probability drops marginally from 0.254 to 0.173 in the moderate case and from 0.598 to 0.355 in the heavy case. Thus, persistent unobserved heterogeneity contributes significantly to time dependence in both types of marijuana consumption, by 32 percent for moderate consumption levels and by just over 40 percent for heavy use.

In row three, we remove persistence in the time-varying utility shocks by setting $\rho = 0$ (in addition to setting $\sigma = 0$). The predicted persistence further drops to 0.156 for the moderate case and to 0.255 for the intense case. This source of persistence contributes about 7 percent to the overall persistence for both moderate levels of marijuana use and 17 percent for heavy levels. Finally, in the last row, we also remove the effect of time-invariant observed characteristics and the time trend by setting $\beta = \kappa_1 = \kappa_2 = 0$. This further reduce the persistence from 0.156 to 0.118 in the moderate case and from 0.255 to 0.196 in the intense case. The remaining persistence (47 percent of the total for moderate use and 33 percent of the total for intense use) is due to a causal or addictive effect of using marijuana in the previous period.

That is, most of the overall persistence in moderate consumption is due to structural state dependence (this result also applies when we consider consumption as a binary outcome) while for heavy consumption, most of the persistence is due to individual heterogeneity.

5 Conclusions

In this paper we provide new evidence on the persistence of marijuana use among American youth. This topic is important for many reason, one being the fact that marijuana consumption among teenagers is inversely related to many successful future labor market outcomes. It is perhaps more important than ever given the recent legalization of recreational marijuana use in many jurisdictions. Moreover, according to 2018 results on monitoring the future from the National Institute on Drug Abuse, marijuana use were at historic highs in 2018, both among college and non-college peers.

The previous literature on persistence of marijuana consumption is limited. A notable exception is Deza (2015) who estimate a dynamic discrete choice model of alcohol, marijuana and hard drugs use and focus on the state dependence in these, as well as dependence across different drugs. While our paper share many features with Deza (2015), there are also important differences. Unlike her, we allow for persistence in the utility shocks, in addition to persistence generated from time-invariant unobserved heterogeneity and pure or causal state dependence. Further, we specify the dynamics in marijuana use in a more flexible way and do not limit it to the inclusion of a one-period lag. Perhaps most importantly, in the second part of the paper, we distinguish between different intensity levels of marijuana consumption. Instead of using a binary outcome (used or not), we code moderate use (consumption during 1-10 days last month) separately from heavy use (consumption during 10 days or more last month). We show that moderate consumption is transitory and less persistent than heavy use. A significant fraction in the data (16.4 percent) of moderate users transit to heavy use in the next period while an even larger share (49.7 percent) stop using marijuana next period.

The estimated average partial effects show that previous consumption significantly increase the probability of current consumption. We show that these effects exist for all consumption levels but are severely exaggerated in models that ignores persistence in utility shocks and restricts the form of dynamics. However, even in the most general model specifications, the partial effects suggest that the probability of consuming marijuana now increase by a factor of 1.5 when we change the status of previous consumption from none to moderate or heavy. This finding is robust towards aggregation of marijuana consumption.

We also disaggregate overall persistence into four components and show the relative contribution of each. The results show that persistent unobserved heterogeneity plays a large role in persistence of heavy marijuana consumption (40 percent of overall persistence is due to unobserved heterogeneity) and less so for moderate use (32 percent is due to unobserved heterogeneity). Persistence in time-varying random shocks also play a significant role and its importance is similar that observed for persistent observed individual characteristics. Finally, true or causal state dependence is important for both intensity levels, but more so for moderate consumption (47 and 33 percent, respectively).

The results for moderate use are similar to those obtained in the binary case where there is no distinction between occasional and intense consumption. These results are also similar to those found in Deza (2015). However, by ignoring the possibility that structural persistence is a function of the level of consumption, the role of causal state dependence may be exaggerated. This in turn may lead to misguided policy recommendations as the risk of addictive behavior may be overstated.

We believe the framework and results provided in this paper will serve as a catalyst for further work in this important area of economics and health. For example, we have restricted the state dependence to be constant across individuals. It would be interesting to investigate if there are differences in persistence between males and females as well as across racial groups. Moreover, we have not considered the consumption of alcohol and cigarettes in this paper but this could be an interesting avenue for future research, building also on the work of Deza (2015). In a companion paper where we estimate the effect of marijuana use on educational attainment, Davalloo and Hansen (2022), we find that age of marijuana initiation is and important determinant of the effect. It may also impact the persistence of marijuana use. These are all topics for extensions of this paper that we plan to pursue in the near future.

References

Belzil, C. & Hansen, J. (2020). The evolution of the US family income-schooling relationship and educational selectivity. *Journal of Applied Econometrics*, 35, 841–859. https://doi.org/10.1002/jae.2785

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

Carro, J. M., & Traferri, A. (2014). State dependence and heterogeneity in health using a bias corrected fixed effects estimator. *Journal of Applied Econometrics*, 29, 181–207.

Cerdá, M., Moffitt, T.E., Meier, M.H., Harrington, H., Houts, R., Ramrakha, S., Hogan, S., Poulton, R. & Caspi, A. (2016). Persistent Cannabis Dependence and Alcohol Dependence Represent Risks for Midlife Economic and Social Problems: A Longitudinal Cohort Study. *Clinical Psychological Science*, 4, 1028–1046.

Chan, G. C. K., Becker, D., Butterworth, P., Hines, L., Coffey, C., Hall, W. & Patton, G. (2021) Young-adult compared to adolescent onset of regular cannabis use: A 20-year prospective cohort study of later consequences. *Drug and Alcohol Review*, 40, 627-636.

Davalloo, G. & Hansen, J. (2022). Heterogeneous Effects of Marijuana Use on Educational Attainment. Working paper, Concordia University.

Deza, M. (2015). Is there a stepping stone effect in drug use? Separating state dependence from unobserved heterogeneity within and between illicit drugs. *Journal of Econometrics*, 184, 193-207.

Fergusson, D.M. & Boden, J.M. (2008). Cannabis use and later life outcomes. *Addiction*, 103, 969-976.

Hansen, J. & Lofstrom, M. (2009). The Dynamics of Immigrant Welfare and Labor Market Behavior. *Journal of Population Economics*, 22, 941-970.

Hall, W. & Degenhardt, L. (2009). Adverse health effects of non-medical cannabis use. *Lancet*, 374, 1383-1391.

Heckman, J. (1981). Heterogeneity and state dependence. in: Rosen, S. (ed.), *Studies in Labor Markets*, University of Chicago Press, Chicago, 91-140.

Honore, B.E., Muris, C. & Weidner, M. (2021) Dynamic Ordered Panel Logit Models. Working paper, Princeton University.

Hyslop, D. (1999). State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women. *Econometrica*, 67, 1255-1294.

Kueng, L. & Yakovlev, E. (2014). How Persistent Are Consumption Habits? Micro-Evidence From Russian Men. NBER working paper 20298.

Lee, L.-F. (1997). Simulated maximum likelihood estimation of dynamic discrete choice statistical models: Some Monte Carlo results. *Journal of Econometrics*, 82, 1-35.

Lev-ran, S. Roerecke, M., Le Foll, B., George, T.P., McKenzie, K. & Rehm, J. (2014). The association between cannabis use and depression: a systematic review and meta-analysis of longitudinal studies. *Psychological Medicine*, 44, 797 - 810.

Liu, X., Kagel, J. & Lee, L.-F. (2012). Dynamic discrete choice models with lagged social interactions: with an application to a signaling game experiment. *Journal* *of Applied Econometrics*, 27, 1037-1058.

McCaffrey, D. F., Pacula, R. L., Han, B., & Ellickson, P. (2010). Marijuana use and high school dropout: The influence of unobservables. *Health Economics*, 19, 1281–1299.

Meier, M. H., Caspi, A., Ambler, A., Harrington, H., & Houts, R. (2012). Persistent cannabis users show neuropsychological decline from childhood to midlife. PNAS, 109, E2657–E2664.

Schmidt, L., Weisner, C. & Wiley, J. (1998). Substance abuse and the course of welfare dependency. *American Journal of Public Health*, 88, 1616–1622.

Van Ours, J. C., & Williams, J. (2009). Why parents worry: Initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics*, 28, 132–142. https://doi.org/10.1016/j.jhealeco.2008.09.001

Wang, Y. (2014). Dynamic Implications of Subjective Expectations: Evidence from Adult Smokers. *American Economic Journal: Applied Economics*, 6, 1-37.

Williams, J., & Van Ours, J. C. (2020). Early cannabis use and school to work transition of young men. *Health Economics*, 29, 1148-1160. https://doi.org/10.1002/hec.4125.

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

Table 1: Proportion of respondents using marijuana, by age.

Age	Used marijuana	Number of individuals
13	0.037	1,204
14	0.091	1,204
15	0.154	1,176
16	0.183	1,142
17	0.204	1,103
18	0.218	1,064
19	0.196	1,024
20	0.200	977
21	0.180	937
22	0.186	916
23	0.160	883
24	0.161	859
25	0.165	843
26	0.162	832

Table 2: Transition matrix

	Used marijuana in year t	
Used marijuana in year t-1	Yes	No
Yes	0.630	0.370
No	0.085	0.915

Note: Row percentages.

Table 3: Sample means, by marijuana use

	Never used	Used at least once
Male	0.49	0.55
Black	0.14	0.13
Hispanic	0.11	0.13
Intact family	0.66	0.57
Family income	$66,191	$65,429
Mother - high school graduate	0.33	0.34
Mother - attend college	0.53	0.53
AFQT	170.9	172.0
Mother's age at birth	26.4	26.2
Urban	0.71	0.75
Number of siblings	2.5	2.4
Peers	0.08	-0.08
Number of individuals	598	606

Note: Family income is expressed in year 2000 dollars.

Table 4: Selected estimates from binary Probits.

	Model 1	Model 2	Model 3
Used marijuana in (t-1)	1.691	0.976	0.732
	(0.032)	(0.045)	(0.053)
σ	-	0.851	0.414
		(0.040)	(0.032)
ρ	-	-	0.220
			(0.083)
Male	0.113	0.157	0.133
	(0.028)	(0.063)	(0.050)
Intact family	-0.113	-0.244	-0.148
	(0.024)	(0.068)	(0.058)
AFQT	0.047	0.074	0.062
	(0.015)	(0.035)	(0.033)
Peers	-0.071	-0.115	-0.076
	(0.011)	(0.034)	(0.025)
AIC	9,422	8,887	8,771
LogL	-4,695	-4,427	-4,368

Note: Standard errors in parentheses. AIC is the Akaike Information Criteria. The dynamics of marijuana use in Models 1 and 2 are assumed to follow a first-order Markov structure. In Model 3, the dynamics is generalized to incorporate use prior to last year. Models 2 and 3 were estimated using simulated Maximum Likelihood with 100 simulation draws.

Table 5: Average partial effects from binary Probits.

	Model 1	Model 2	Model 3
Used marijuana in (t-1)	0.473	0.192	0.129
	(0.009)	(0.013)	(0.014)
Male	0.019	0.021	0.017
	(0.005)	(0.008)	(0.006)
Intact family	-0.018	-0.032	-0.018
	(0.004)	(0.009)	(0.008)
AFQT	0.008	0.010	0.007
	(0.002)	(0.005)	(0.004)
Peers	-0.012	-0.016	-0.010
	(0.002)	(0.004)	(0.003)

Note: Standard errors in parentheses. AIC is the Akaike Information Criteria. The dynamics of marijuana use in Models 1 and 2 are assumed to follow a first-order Markov structure. In Model 3, the dynamics is generalized to incorporate use prior to last year. Models 2 and 3 were estimated using simulated Maximum Likelihood with 100 simulation draws. A parametric bootstrap with 100 draws was used to estimate the standard errors.

Table 6: Predicted transition matrices

	Used marijuana in year t	
Used marijuana in year t-1	Yes	No
Yes	0.661	0.339
	(0.159)	(0.159)
No	0.099	0.901
	(0.054)	(0.054)

Note: Average transition probabilities from simulation of outcomes using estimates from Model 3 in Table 4 (the Polya model). Standard errors in parentheses. A parametric bootstrap with 100 draws was used to estimate the standard errors.

Table 7: Sources of persistence

	Polya
(1) Predicted persistence	0.661
(2) Removing time-invariant unobserved heterogeneity	0.567
Proportion of total persistene - (2)/(1)	*0.857*
(3) Removing time-varying unobserved characteristics and (2)	0.449
Proportion of total persistene - (3)/(1)	*0.679*
(4) Removing observed characteristics, time trend and (3)	0.345
Proportion of total persistene - (4)/(1)	*0.522*

Note: The entries are derived using estimates from Model 3 in Table 4 and show $Pr\left(y_{i,t}=1|y_{i,t-1}=1\right)$. In (2), we set $\sigma_u = 0$ and in (3), we set $\sigma_u = 0; \rho = 0$. Finally, in (4), we set $\sigma_u = 0; \rho = 0; \beta = 0; \kappa_1 = 0; \kappa_2 = 0$.

Table 8: Proportion of respondents using marijuana, by age.

Age	Did not use marijuana	Used marijuana less than 10 days	Used marijuana 10 days or more	Number of individuals
13	0.963	0.027	0.010	1,204
14	0.909	0.073	0.018	1,204
15	0.846	0.107	0.047	1,176
16	0.817	0.122	0.061	1,142
17	0.796	0.119	0.085	1,103
18	0.782	0.116	0.102	1,064
19	0.804	0.104	0.093	1,024
20	0.800	0.107	0.092	977
21	0.820	0.099	0.081	937
22	0.814	0.094	0.092	916
23	0.840	0.079	0.080	883
24	0.839	0.079	0.081	859
25	0.835	0.077	0.088	843
26	0.838	0.064	0.099	832

Table 9: Transition matrix

	Days of marijuana use last month in year t		
	0	1-9	10 or more
Days of marijuana use last month in year t-1			
0	0.915	0.064	0.021
1-9	0.497	0.339	0.164
10 or more	0.198	0.166	0.635

Note: Row percentages.

Table 10: Selected estimates from an ordered Probit polya model

	Estimate	Standard error
Marijuana, 1-9 days (t-1)	0.432	0.047
Marijuana, 10+ days (t-1)	0.786	0.051
Male	0.143	0.047
Intact family	-0.183	0.054
AFQT	0.056	0.031
Peers	-0.088	0.026
σ	0.569	0.060
ρ	0.300	0.025
θ_1	1.735	0.182
θ_2	2.556	0.186
LogL	-5,623	

Note: The specification included additional observed characteristics (the same list as in Table 4). The remaining parameter estimates and standard errors are presented in Table A2 in Appendix.

Table 11: Average partial effects from selected variables on the probability of moderate and heavy marijuana consumption.

	Moderate	Heavy
$Pr\left(y^m_{i,t}=1\vert y^m_{i,t-1}=1\right)-Pr\left(y^m_{i,t}=1\vert y^n_{i,t-1}=1\right)$	0.046 (0.001)	-
$Pr\left(y^m_{i,t}=1\vert y^m_{i,t-1}=1\right)-Pr\left(y^m_{i,t}=1\vert y^h_{i,t-1}=1\right)$	-0.051 (0.001)	-
$Pr\left(y^h_{i,t}=1\vert y^h_{i,t-1}=1\right)-Pr\left(y^h_{i,t}=1\vert y^n_{i,t-1}=1\right)$	-	0.043 (0.001)
$Pr\left(y^h_{i,t}=1\vert y^h_{i,t-1}=1\right)-Pr\left(y^h_{i,t}=1\vert y^m_{i,t-1}=1\right)$	-	0.027 (0.001)
Male	0.013 (0.001)	0.004 (0.001)
Intact family	-0.016 (0.001)	-0.005 (0.001)
AFQT	0.003 (0.001)	0.001 (0.001)
Peers	-0.005 (0.001)	-0.001 (0.001)

Note: A parametric bootstrap with 100 draws was used to estimate the standard errors of the average partial effects.

Table 12: Model fit: Transition matrix

	Days of marijuana use last month in year t		
	0	1-9	10 or more
Days of marijuana use last month in year t-1			
0	0.923	0.060	0.017
	(0.003)	(0.002)	(0.001)
1-9	0.551	0.254	0.195
	(0.002)	(0.003)	(0.003)
10 or more	0.183	0.219	0.598
	(0.004)	(0.003)	(0.006)

Note: Row percentages.

Table 13: Sources of persistence

	Persistence	
	Moderate	Heavy
(1) Predicted persistence	0.254	0.598
(2) Removing time-invariant unobserved heterogeneity	0.173	0.355
Proportion of total persistene - (2)/(1)	*0.680*	*0.594*
(3) Removing time-varying unobserved characteristics and (2)	0.156	0.255
Proportion of total persistene - (3)/(1)	*0.615*	*0.427*
(4) Removing observed characteristics, time trend and (3)	0.118	0.196
Proportion of total persistene - (4)/(1)	*0.467*	*0.328*

Note: The entries are derived using estimates from the model presented in Table 10 and show $Pr\left(y^j_{i,t} = 1 | y^j_{i,t-1} = 1\right)$, $j = Moderate, Heavy$. A parametric bootstrap with 100 draws was used to estimate the standard errors. In (2), we set $\sigma_u = 0$ and in (3), we set $\sigma_u = 0; \rho = 0$. Finally, in (4), we set $\sigma_u = 0; \rho = 0; \beta = 0; \kappa_1 = 0; \kappa_2 = 0$.

Appendix

Table A1: Estimates from binary Probits.

	Model 1	Model 2	Model 3
Black	-0.079	-0.153	-0.054
	(0.064)	(0.103)	(0.074)
Hispanic	0.004	0.046	0.029
	(0.066)	(0.071)	(0.079)
Family income	0.001	0.004	0.002
	(0.003)	(0.006)	(0.005)
Mother High School	0.049	0.026	0.033
	(0.049)	(0.092)	(0.076)
Mother College	0.015	-0.015	-0.005
	(0.052)	(0.092)	(0.099)
Mother's age	-0.001	0.0003	-0.001
	(0.003)	(0.007)	(0.005)
Urban	0.037	0.095	0.053
	(0.037)	(0.029)	(0.049)
Siblings	-0.049	-0.078	-0.052
	(0.021)	(0.031)	(0.029)
$(t - t_0)$	0.077	0.180	0.088
	(0.017)	(0.019)	(0.029)
$(t - t_0)^2$	-0.005	-0.011	-0.008
	(0.001)	(0.001)	(0.002)
Constant	-1.508	-1.970	-1.571
	(0.159)	(0.235)	(0.198)

Note: Standard errors in parentheses. The remaining parameters and model descriptions are available in Table 4 together with likelihood values and AIC.

Table A2: Estimates from an ordered Probit polya model

	Estimate	Standard error
Black	-0.056	0.082
Hispanic	0.047	0.075
Family income	0.001	0.005
Mother High School	0.031	0.060
Monther College	-0.006	0.056
Mother's age at birth	-0.0002	0.005
Urban	0.082	0.054
Siblings	-0.050	0.026
$(t - t_0)$	0.124	0.023
$(t - t_0)^2$	-0.010	0.001

Note: The remaining parameters and model descriptions are available in Table 10 together with likelihood values and AIC.

Generation of truncated random variables for the simulated likelihood function

Binary outcomes

In order to derive the likelihood function in equation (4), we need to generate random variables ($e_{i,t}$) from truncated standard normal distributions on $[L_{i,t}, U_{i,t}]$. This can be done by transformations of uniformly distributed random variables, $u_{i,t} \sim U[0,1]$. Specifically, for each independent simulation run (j), $e_{i,t}$ can be recursively generated as follows (see also Lee (1997)).

1. Draw μ_i from a standard normal distribution.

2. For the first period,

   (a) Calculate $d_{i,1} = \Psi_{i,1} + \sigma\mu_i$ (assuming the following initial conditions $\varepsilon_{i,0} = 0$ and $y_{i,0} = 0$ for all individuals)

   (b) Calculate $a_{i,1} = \Phi(d_{i,1}) * I(y_{i,1} = 1) + \Phi(-d_{i,1}) * I(y_{i,1} = 0)$

   (c) Calculate $b^0_{i,1} = u_{i,1} * \Phi(-d_{i,1})$

   (d) Calculate $b^1_{i,1} = \Phi(-d_{i,1}) + u_{i,1} * \Phi(d_{i,1})$

   (e) Calculate $e_{i,1} = \Phi^{-1}\left(b^0_{i,1}\right) * I(y_{i,1} = 0) + \Phi^{-1}\left(b^1_{i,1}\right) * I(y_{i,1} = 1)$

   (f) Obtain $\varepsilon_{i,1} = e_{i,1}$

3. For $t > 1$,

   (a) Calculate $d_{i,t} = \Psi_{i,t} + \gamma \sum_{j=1}^{t} \delta^{j-1} y_{i,t-j} + \sigma\mu_i + \rho\varepsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ is drawn from a standard normal distribution

   (b) Calculate $a_{i,t} = \Phi(d_{i,t}) * I(y_{i,t} = 1) + \Phi(-d_{i,t}) * I(y_{i,t} = 0)$

(c) Calculate $b^0_{i,t} = u_{i,t} * \Phi(-d_{i,t})$

(d) Calculate $b^1_{i,t} = \Phi(-d_{i,t}) + u_{i,t} * \Phi(d_{i,t})$

(e) Calculate $e_{i,t} = \Phi^{-1}\left(b^0_{i,t}\right) * I(y_{i,t} = 1) + \Phi^{-1}\left(b^1_{i,t}\right) * I(y_{i,t} = 0)$

(f) Obtain $\varepsilon_{i,t} = e_{i,t} + \rho\varepsilon_{i,t-1}$

This is done m times. The simulated likelihood is then

$\mathcal{L} = \sum_{i=1}^{n} ln\left\{\frac{1}{m}\sum_{j=1}^{m}\prod_{t=1}^{T_i} a_{i,t}\right\}$

Asymptotic properties of this estimator are discussed in Lee (1997) as well as in the references in that paper.

Ordered outcomes

The simulated likelihood function for the dynamic ordered Probit proceeds in a similar fashion but modified to accomodate the ternary nature of our outcomes. Specifically, for each independent simulation run (j), $e_{i,t}$ can be recursively generated as follows:

1. Draw μ_i from a standard normal distribution.

2. For the first period,

   (a) Calculate $d_{i,1} = \Psi_{i,1} + \sigma\mu_i$ (assuming the following initial conditions $\varepsilon_{i,0} = 0$ and $c_{i,0} = 0$ for all individuals)

   (b) Calculate $a_{i,1} = \Phi(\theta_1 - d_{i,1}) * I(c_{i,1} = 0) + [\Phi(\theta_2 - d_{i,1}) - \Phi(\theta_1 - d_{i,1})] * I(c_{i,1} = 1) + [1 - \Phi(\theta_2 - d_{i,1})] * I(c_{i,1} = 2)$

   (c) Calculate $b^0_{i,1} = u_{i,1} * \Phi(\theta_1 - d_{i,1})$

   (d) Calculate $b^1_{i,1} = \Phi(\theta_1 - d_{i,1}) + u_{i,1} * [\Phi(\theta_2 - d_{i,1}) - \Phi(\theta_1 - d_{i,1})]$

   (e) Calculate $b^2_{i,1} = \Phi(\theta_2 - d_{i,1}) + u_{i,1} * [1 - \Phi(\theta_2 - d_{i,1})]$

(f) Calculate $e_{i,1} = \Phi^{-1}\left(b_{i,1}^{0}\right) * I\left(c_{i,1} = 0\right) + \Phi^{-1}\left(b_{i,1}^{1}\right) * I\left(c_{i,1} = 1\right) + \Phi^{-1}\left(b_{i,1}^{2}\right) * I\left(c_{i,1} = 2\right)$

(g) Obtain $\varepsilon_{i,1} = e_{i,1}$

3. For $t > 1$,

   (a) Calculate $d_{i,t} = \Psi_{i,t} + \gamma_1 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 1\right) + \gamma_2 \sum_{j=1}^{t} \delta^{j-1} 1\left(c_{i,t-1} = 2\right) + \sigma\mu_i + \rho\varepsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ is drawn from a standard normal distribution

   (b) Calculate $a_{i,t} = \Phi\left(\theta_1 - d_{i,t}\right) * I\left(c_{i,t} = 0\right) + \left[\Phi\left(\theta_2 - d_{i,t}\right) - \Phi\left(\theta_1 - d_{i,t}\right)\right] * I\left(c_{i,t} = 1\right) + \left[1 - \Phi\left(\theta_2 - d_{i,t}\right)\right] * I\left(c_{i,t} = 2\right)$

   (c) Calculate $b_{i,t}^{0} = u_{i,t} * \Phi\left(\theta_1 - d_{i,t}\right)$

   (d) Calculate $b_{i,t}^{1} = \Phi\left(\theta_1 - d_{i,t}\right) + u_{i,t} * \left[\Phi\left(\theta_2 - d_{i,t}\right) - \Phi\left(\theta_1 - d_{i,t}\right)\right]$

   (e) Calculate $b_{i,t}^{2} = \Phi\left(\theta_2 - d_{i,t}\right) + u_{i,t} * \left[1 - \Phi\left(\theta_2 - d_{i,t}\right)\right]$

   (f) Calculate $e_{i,t} = \Phi^{-1}\left(b_{i,t}^{0}\right) * I\left(c_{i,t} = 0\right) + \Phi^{-1}\left(b_{i,t}^{1}\right) * I\left(c_{i,t} = 1\right) + \Phi^{-1}\left(b_{i,t}^{2}\right) * I\left(c_{i,t} = 2\right)$

   (g) Obtain $\varepsilon_{i,t} = e_{i,t} + \rho\varepsilon_{i,t-1}$

Similar to the binary case, this is done m times and the simulated likelihood is

$\mathcal{L} = \sum_{i=1}^{n} ln\left\{\frac{1}{m}\sum_{j=1}^{m}\prod_{t=1}^{T_i} a_{i,t}\right\}$

Asymptotic properties of this estimator are discussed in Lee (1997) as well as in the references in that paper.