Marijuana on Main Street?
Estimating Demand in Markets with Limited Access

Liana Jacobi and Michelle Sovinsky*

August 7 2013

Abstract

Marijuana is the most common illicit drug with vocal advocates for legalization. Among other things, legalization would increase access and remove the stigma of illegal behavior. Our model of use disentangles the role of access from preferences and shows that selection into access is not random. We find that non-selection-corrected demand estimates are biased resulting in incorrect policy conclusions. Our results show the probability of underage use would increase by 38% and more than double in most age groups under legalization. Tax policies are effective at curbing use where over $8 billion in annual tax revenue could be realized.

JEL Classification: L15, K42, H27

Keywords: legalization, drugs, discrete choice models, limited information, selection correction

Marijuana is the most widely used illicit drug in the world (ONDCP, 2004). According to the United Nations Office of Drugs and Crimes (2012), there are 119 to 224 million users worldwide. While the nature of the market makes it difficult to determine total sales with certainty, estimates indicate sales in the United States are around $150 billion per year (Miron, 2005). Despite the attempts to regulate use, in nearly every country, the market for illicit drugs remains pervasive.

The marijuana market has the most vocal advocates for legalization of all illicit drugs. Within Europe: Germany, the Netherlands, Portugal, Spain and Switzerland all currently exhibit liberal attitudes of law enforcement towards marijuana possession. The United States has a more punitive system, but in January 2013 recreational use of marijuana became legal in Washington and Colorado.[1] In Australia, there have been many campaigns in the

---

[0] *Liana Jacobi is at the University of Melbourne, Department of Economics, 111 Barry Street Carlton VIC 3053, Australia, ljacobi@unimelb.edu.au; Michelle Sovinsky is at the the University of Zurich, Department of Economics, Blumlisalpstrasse 10, CH-8006 Zurich, Switzerland and CEPR, michelle.sovinsky@gmail.com. Corresponding author is Sovinsky. We are grateful to Peter Arcidiacono, Eve Caroli, Karim Chaib, Sofronis Clerides, Rosa Ferrer, T. Head, Carlos Noton, Alison Ritter, Terry Sovinsky, and seminar participants at Adelaide, Bocconi, Duke, Rotman School of Business, Cyprus, Dauphine (Paris), Monash, Sydney, University Technology Sydney, Virginia, Warwick, the Barcelona Summer Forum, the Cannabis Policy Conference (Melbourne), the Conference for Mental Health and Well-Being (Melbourne), and the IIOC (Washington DC) meetings for helpful comments and suggestions.

[1] Adults in Washington and Colorado are allowed to grow one plant and possess up to one ounce for their

larger cities to legalize cannabis. Indeed, for the past 30 years there has been a debate regarding marijuana legalization in many countries.[2] Those in favor of legalization cite the harsh consequences a criminal record can have for young users who are otherwise law-abiding citizens, the costs of black-market violence, the exposure to harder drugs from dealer interactions, the high expenditures on enforcement, and the foregone sales tax revenues. Those opposed are concerned about the impact on health outcomes and that legalization could result in lower prices, hence generating higher use. This is of particular concern if use among young adults increases and marijuana usage serves as a "gateway" to subsequent consumption of other harder drugs (Bretteville-Jensen and Jacobi, 2011; DeSimone, 1998; and Van Ours, 2003).

Much of the discussion surrounding marijuana drug policy is concerned with the following questions. First, by how much would the prevalence and intensity of use rise under legalization? Second, to what extent would at risk groups (such as youth) be impacted by legalization? Finally, could government policies (such as taxation) be effective in curbing use? In this paper, we provide a methodology for examining the consequences of legalizing illicit drugs, which helps lead us to answers to these questions.

During the last two decades there have been many empirical studies that assess the impact of decriminalization on marijuana use. These include Caulkins, et. al. (2011), Clements and Zhao (2009), Donohue, Ewing, and Peloquin (2011), Miron and Zwiebel (1995), Pacula, et. al. (2000), Pacula, et. al. (2010), Pudney (2010), and Williams, van Ours, and Grossman (2011).[3] However, decriminalization and legalization differ in significant ways. The first important way concerns limited accessibility. Given that illicit drugs are not as easy to find as legal products, one can argue that non-users have very little information about how to get marijuana, which is the first step to becoming a user. Under decriminalization it is still necessary to seek out suppliers in order to purchase the drug. If marijuana were legalized, purchasing it would be as difficult as purchasing cigarettes or alcohol. Second, while decriminalization removes criminal penalties, using the drug is still illegal. In fact, in the Australian National Drug Household Survey, a significant fraction of non-users report

own use. Selling or growing for commercial use is not legal. Marijuana use is already decriminalized in some states, where possession is an infraction, the lowest level of offence under state law. For example, an adult in California caught with an ounce of marijuana will get a $100 ticket but not a criminal conviction.

[2] Pacula, et. al. (2010) provides a literature review.

[3] There is also a large literature on drug policies and the effect of decriminalization or enforcement on crime; see, for example, Adda, et. al. (2011) and Sickles and Taubman (1991).

not using marijuana because it is illegal. Legalization would obviously remove this stigma (and cost) associated with illegal behavior, which may result in use among some current non-users. The third way in which decriminalization and legalization differ concerns the impact on dealers. Decriminalization makes it less costly for potential users in that they face a fine for using the drug instead of the harsher cost of a criminal punishment. In contrast, selling the drug is still illegal and hence dealers, should they be arrested, incur the same penalties regardless of the decriminalization status of the state. In other words, decriminalization does not impact the costs (broadly defined to include the risk of criminal prosecution) faced by dealers, while legalization eliminates the risk of arrest leading to lower costs. For these reasons, models that focus on the impact of decriminalization will not provide us with answers to what will happen to use under legalization.

This paper provides the first approach to modeling and estimating the impact of legalization on use. To do so we explicitly consider the role played by accessibility in use, the impact of illegal actions on utility, as well as the impact on the supply side. We present a model of buyer behavior that includes the impact of illegal behavior on utility as well as the impact of limited accessibility (either knowing where to buy or being offered an illicit drug) on using marijuana. We apply the model to data collected in the Australian National Drug Strategy Household Survey.[4] These data are particularly suited for our purposes as they contain information both on use and also on access and enable us to identify the preference parameters on marijuana use. For example, we obtain estimates for price elasticities of demand (for an illicit good) taking into account selection into access. Modeling both of these effects is particularly important for drawing correct inferences about choices that individuals would make under a policy of legalization, where the accessibility issue would essentially disappear. We find that predictions based on a model that doesn't consider selection are biased due to ignoring the important role that selection based on observables and unobservables plays in the context of marijuana use and more generally in the use of illicit drugs.

Our modelling framework also directly addresses an issue that is prevalent in studies of illicit markets: the fact that prices are not observed for each purchase. To do so we use extra individual-level data on the type of marijuana used (i.e., head, hydro) combined with market-level pricing data to obtain an implied price faced by users and non-users. This

[4] Several studies use these data to eamine issues related to marijuana, such as Damrongplasit, et. al. (2010) and Williams (2004).

allows us to estimate a model with individual prices while not observing these in the data.

We use the demand side estimates to conduct counterfactuals on how use would change under legalization, how effective government policies would be at curbing use, and what tax revenues could be raised under legalization. We also consider differences across age groups (including teenagers) and conduct counterfactuals of how much taxes would need to be imposed to return the probability of underage use to what it was before legalization (at the individual level). The counterfactual analysis is implemented under different post-legalization prices to allow for different supply side scenarios.

We find that selection into who has access to marijuana is not random, and the results suggest estimates of the demand curve will be biased unless selection is explicitly considered. Our results indicate that if marijuana were legalized and accessibility were not an issue the probability of use would more than double to 28.7%. In addition, the average probability of underage use would increase by 38% (to 31.8% from 23.1%). For a population the size of the US our results indicate that a policy of marijuana legalization would raise a minimum amount between $452 million and $8 billion annually depending on the tax scheme implemented after accounting for lost sales to the black market. We find that prices would need to be more than ten-fold higher than current levels in order to keep the frequency of post-legalization underage use the same as pre-legalization use even if underage users would still face the same restrictions as they face for alcohol use. Increasing prices by ten-fold is not feasible given that we would expect most users to resort to the black market. Hence, our results indicate that, while teens respond to prices, an environment where use is legalized will see an increase in the probability of use of at least 5% and at most 67% among teens.

The previous literature on decriminalization already mentioned is not concerned with the impact of limited access on consumption decisions. In this sense, the approach presented in our paper is conceptually more closely related to the empirical IO literature that examines markets with limited consumer information. These include papers by Ching, Erdem and Keane (2009), Clerides and Courty (2010), Eliaz and Spiegler (2011), Kim, Albuquerque, and Bronnenberg (2010), and Sovinsky Goeree (2008). There is also a small but growing related literature that addresses sample selection issues in empirical IO including Eizenberg (2011) and Veugelers and Cassiman (2005).

The paper is structured as follows. Section 1 gives an overview of that data and legal policies in Australia. Sections 2 and 3 outline the model and the estimation technique. We discuss our parameter estimates in Section 4. We present results of counterfactual policy experiments in Section 5. We examine the robustness of our results to alternative specifications in Section 6. Finally, we conclude and discuss directions for future work in Section 7.

1 Data

Cannabis comes in a variety of forms and potency levels. The herbal form consists of the dried flowering tops, leaves and stalks of the plant. The resinous form consists of the resin secreted from the plant and resin oil. In this paper we focus on the most commonly used forms of cannabis: the leaf of the plant, the flowering tops (or head) of the plant, and a high potency form selectively bred from certain species (sinsemilla, called skunk). The leaf, head, and skunk are collectively known as marijuana.[5]

The major psychoactive chemical compound in marijuana is delta-9-tetrahydrocannabinol (or THC). The amount of THC absorbed by marijuana users differs according to the part of the plant that is used, the way the plant is cultivated, and the method used to imbibe the drug. On average marijuana contains about 5% THC, where the flowering tops contain the highest concentration followed by the leaves (Adams and Martin, 1996). Cannabis that is grown hydroponically (hydro), indoors under artificial light with nutrient baths, typically has higher concentrations of THC than naturally grown marijuana (Poulsen and Sutherland, 2000). Given that the forms of marijuana vary in THC content, and users may select the forms based on THC content, in the model we include a variable to capture the level of THC, which can be thought of as the "quality" of the marijuana product.

In Australia the use of cannabis for any purpose is illegal, however, all states/territories have introduced legislation to allow police to deal differently with minor offenses. Four jurisdictions (South Australia (SA), Northern Territory (NT), Australia Capital Territory (ACT), and Western Australia (WA)) have decriminalized the possession of small quantities of cannabis via the introduction of infringement schemes. Under an infringement scheme individuals which are found to have violated the law with a minor marijuana offence are fined but are not jailed. Infringement schemes were introduced at different times across the states: SA was the first to implement them in 1987, followed by NT in 1992 and ACT in 1996. In 2004 WA moved to this system. What constitutes a minor offense and the fine varies by

[5] We do not consider hashish (the resin or resin oil of the plant) as these forms are much harder to obtain and have a much higher level of the psychoactive component.

state.[6] In other states and territories (Tasmania (TAS), Victoria (VIC), New South Wales (NSW), and Queensland (QLD)) possession of any amount of cannabis is a criminal offence, and individuals may be jailed for possession of any quantity.[7] We construct two measures of the degree of decriminalization. These include whether the state uses an infringement scheme and the maximum number of grams for which possession is a minor offense.

1.1 Cannabis Use and Access

We use data from an individual-level cross-section survey called the Australian National Drug Strategy Household Survey (NDSHS). The NDSHS was designed to determine the extent of drug use among the non-institutionalized civilian Australian population aged 14 and older.[8] About 20,000 (different) individuals are surveyed every 2 or 3 years from all Australian states/territories. We use data from three waves: 2001, 2004, and 2007. These data are particularly useful as they not only contain demographic, market, and illicit drug use information, but they also contain a number of variables on accessibility to marijuana. These latter questions are crucial in order to estimate our model.

Table 1 presents descriptive statistics of our sample. We restrict the data to individuals aged between 16 and 60. The average age of a respondent in our sample is just under 40. Approximately 60% are married and 2% of the sample are of Aboriginal descent. About 60% of the sample live in a major city. We construct an indicator variable equal to one if individuals report their health status is good, very good, or excellent. About 56% of individuals report being in good or better health.[9]

The second panel presents information about cannabis use. Nearly half of the population has tried marijuana at least once in their lifetime, where the average age of onset is 19. In

[6] These include possession of small amount of marijuana plant material (i.e., bulbs, leaves)(SA and NT), growing of one plant (SA) or two plants. The quantity considered a minor offence varies by cannabis type (plant versus resin) and range from 100 grams of plant material in SA to 25 grams in ACT.

[7] These jurisdictions have introduced “diversion schemes” where the police may issue a caution of diversion into treatment or education for a minor offence instead of jail time. The number of cautions issued before a criminal conviction varies by jurisdictions. The diversion schemes were introduced at different times: in 1998 in TAS and VIC; in 2000 in NSW, and 2001 in QLD. The state of WA gradually introduced the schemes between 2000 to 2003. Minor cannabis offences only refer to the possession of cannabis, not the possession of a plant. Trafficking and possessions of larger amounts of cannabis are serious offences that incur large monetary fines and long prison sentences.

[8] Respondents were requested to indicate their level of drug use and the responses were sealed so the interviewer did not know their answers.

[9] Our measure of health status is the self-reported answer to “Would you say your health is: 1=excellent; 2=very good; 3=good; 4=fair; 5=poor.”

every year the survey asks "Have you used marijuana in the last 12 months?" In 2001 just over 16% reported using marijuana in the past year, but this declined to around 12% by 2007. Although the rates of marijuana use are considerable, most people who use marijuana do not use on a daily basis. Those that report they use marijuana daily or habitually is around 3%. We classify users as infrequent if they use only quarterly, biannually or annually and as frequent if they use more often. Users are approximately evenly split among these two groups. We should note that hard core drug users are less likely to return the survey or to be available for a telephone survey. Hence, our study will reflect more recreational users.

			Year	
		2001	2004	2007
Demographics				
Male		43%	42%	42%
Age		38	39	40
Aboriginal Descent		2%	2%	2%
Live in City		62%	60%	59%
In Good, Very Good, or Excellent Health		57%	54%	58%
High School Education		16%	15%	14%
Trade Degree		36%	35%	37%
University Degree		22%	25%	28%
Cannabis Use				
Ever Used		44%	45%	46%
Used in Last 12 Months		16%	15%	12%
Use Infrequently (Quarterly, Biannually or Annually)		8%	7%	6%
Use Frequently (Monthly, Weekly or Daily)		8%	8%	6%
Report Use is a Habit		3%	3%	2%
Use Daily		3%	3%	2%
Would Use/UseMore if Legal		6%	7%	6%
Average Age First Used		19	19	19
Access Variables				
Impossible	(all access=0)	9%	13%	14%
No Opportunity to Use	(all access=0)	8%	8%	6%
Fairly Difficult to Obtain	(access1=1)	6%	7%	8%
Fairly Easy to Obtain	(access1 and access2=1)	22%	23%	25%
Very Easy to Obtain	(all access=1)	34%	29%	19%
Offered Cannabis	(all access=1)	28%	26%	22%
Access 1		63%	61%	53%
Access 2		58%	54%	46%
Access 3		42%	37%	29%
Number of Observations		18370	19583	13343

Notes: 48 individuals reported "no opportunity to use" as a reason they had not used but were recorded as using in the last 12 months; 45 individuals reported that cannabis was not available to them but were recorded as using in the last 12 months. We dropped these 93 individuals.

Table 1: Annual Descriptive Statistics

To assess the role the legal status of marijuana plays in the decision to use, we construct a variable that is intended to capture the (dis)utility associated with doing something illegal. It is defined from responses to questions of the form "If marijuana/cannabis were legal to use, would you..." where we set the (dis)taste for illegal behavior variable equal to: 0 if the answer is "Not use it - even if legal and available;" equal to 1 if the answer is "Try it; or Use it as often/more often than now." Some individuals appear to get positive utility from doing something that is illegal. These individuals report they would "use [cannabis] less often than now" if using it were legalized. We set our (dis)taste for illegal behavior variable equal to -1 for these individuals. About 6% of individuals who don't use marijuana (but report they know where to get it) say they would use it if it were legal. Among current users, approximately 13% report they would use marijuana more often than they currently do if it were legalized.

The NDSHS survey also asks questions regarding how accessible marijuana is to the individual, which is particularly suited to the focus of this research. We construct three different measures of accessibility based on the answers to these questions. While not all individuals answered all the questions we have answers to at least one question for each user. If the individual reports that they had the opportunity to use or had been offered the drug in the past 12 months (about 25% of the sample) then they must have had access to the drug, so we set all our accessibility measures to one. Second, they report how difficult it would be to obtain marijuana. If they indicate it is very easy (about 27% of the sample) then we set all accessibility variables to one; or if the response is impossible (about 12% of the sample) then we set all access variables equal to zero. Third, non-users were asked why they didn't use the drug. If they answer it was "too difficult to get" or they had "no opportunity" (about 8% of the non-user sample) then we set all accessibility variables to zero.[10] We examine the robustness of our results to our definition of accessibility by constructing different definitions of access depending on the ease with which they report they could obtain marijuana. Our most broad definition (access 1) indicates about 62% of the sample has access to marijuana; our intermediate measure (access 2) is 53% on average; whereas our most restrictive definition indicates under 40% have access. The results in this paper are presented using the intermediate definition of access (access 2) and robustness of

[10] There were 48 individuals which reported no opportunity to use as a reason they had not used but who where recorded as using in the last 12 months; there were an additional 45 individuals who reported that cannabis was not available to them but who used in the last 12 months. We drop these 93 observations.

the results to the access variable are presented in section 6.

Table 2 provides descriptive statistics by access and use (based on the intermediate definition of access). Cannabis use and access varies with age and is the most prevalent among those in their twenties and thirties. Use declines to under 0.4% for those in their sixties. Males and younger people are more likely to have access and, conditional on having access to use marijuana. Cannabis use varies across states, ranging from 12% in Victoria to over 20% in the Northern Territory. Not surprisingly both use and access are higher in states where marijuana use is decriminalized. If we compute the percentage of users among those with access (as opposed to the percentage of users among the entire population) the percent with access that report using marijuana has a higher mean and lower variance across states.

Demographic Group or State	Percent Used in Last 12 Months	Percent Report Access	Percent With Access that Use	Average State Price	Number of Observations
On Average	15%	53%	27%		51296
Male	18%	59%	31%		21740
Teenager	28%	72%	38%		3275
Age in Twenties	27%	73%	37%		9901
Age in Thirties	16%	59%	27%		13361
Age in Forties	10%	48%	22%		12183
Age over Forties	4%	33%	11%		12576
New South Wales	13%	51%	26%	41.79	13910
Victoria	13%	49%	26%	33.51	10758
Queensland	14%	52%	28%	33.09	9230
Western Australia	19%	61%	32%	42.31	5744
South Australia	15%	58%	27%	41.05	4152
Tasmania	15%	58%	26%	26.08	2290
ACT	14%	52%	27%	28.38	2614
Northern Territory	21%	65%	32%	38.18	2598
Decrimilized	16%	58%	28%	38.90	12743
Not Decrimilized	14%	52%	27%	36.26	38553

Notes: These are based on access variable definition 2.

Table 2: Descriptive Statistics by Use and Access

1.2 Prices

Our market-level pricing data comes from the Australian Bureau of Criminal Intelligence, Illicit Drug Data Reports which are collected during undercover buys. Given that marijuana is an illicit drug there are a few data issues to resolve regarding the prices. First, we do not observe prices in all years due to different state procedures in filling in forms and the frequency of drug arrests of that certain marijuana form. To deal with missings across time we use linear interpolation when we observe the prices in other years. Second, the price per gram is the most frequently reported price, but in some quarters the only price available is the price per ounce.[11] We cannot simply divide the price per ounce by 28 to convert it to grams as quantity discounts are common (Clements 2006). However, assuming price changes occur at the same time with gram and ounce bags, when we observe both the gram and ounce prices we substitute the corresponding price per gram for the time period in which it is missing when the price per ounce is the same in the period where both are reported. Third, some prices are reported in ranges, in which case we use the mid-point of the reported price range. Finally, when skunk prices are not available we use the price per gram for hydro. We deflate the prices using the Federal Reserve Bank of Australia Consumer Price Index for Alcohol and Tobacco where the prices are in real 1998 AU$. These data are reported on a quarterly or semi-annual basis. We construct an annual price per gram measure by averaging over the periods.[12]

		Year	
	2001	2004	2007
Median Market Prices per Gram			
Leaf	30	33	37
Head	30	34	37
Hydro	32	34	38
Individual Use by Type			
Leaf	46%	43%	38%
Head	79%	76%	68%
Hydro	23%	19%	40%
Number of Observations	18370	19583	13343

Notes: These are real prices in 1998$.

Table 3: Prices and Use by Type

The survey contains information about which form of marijuana the user uses and how often. We use these information together with type-specific prices to simulate an individual

[11] A joint contains between 0.5 to 1.5 grams of plant material.

[12] We also considered using pricing data reported in the Illicit Drug Reporting System National Reports. These are self-reported prices from users. Unfortunately they are less believable in that there is virtually no variation in nominal prices across years, states, and quality types: 88% of the observations are either 20 or 25 (with a mean of 23 and standard deviation of 3).

price for person i. The details of how we construct simulated prices for users, as well as for non-users, are discussed in detail in section 3. Table 3 present the percentage of use per type and market prices across years. Notice that it is common to use types in combination (i.e., a bag might contain leaf and head), hence the percentages do not sum to one. Not surprisingly given the higher amount of THC present in hydro, hydro demands a higher price. The lower panel shows that users have moved into using more hydro in the latter year. This is consistent with patterns seen in the rest of the world.[13]

2 Model

Our paper concerns the impact of legalization on marijuana use. Given that illicit drugs are not as easy to find as legal products, one can argue that non-users have less information about how to get marijuana, which is the first step to becoming a user. If marijuana were legalized, purchasing it would be as difficult as purchasing cigarettes or alcohol. Furthermore, legalization would remove the "breaking the law" hindrance, which may result in use among some current non-users.

An individual chooses whether or not to consume marijuana in market m which is defined as a state-year combination.[14] The indirect utility individual i obtains from using marijuana in market m depends on a number of factors including the price the individual pays (p_i), their demographic characteristics (represented by the vector d_i), such as gender, age splines (in young adult, college age, pensioner, etc.), health status, and whether they are of aboriginal descent.[15] Market specific variables can also impact the benefit of consuming marijuana. These are represented by x_m and include the year in which the marijuana was purchased, state-fixed effects, and the proportion of high quality marijuana sold in the market. We also include variables related to legality represented by the vector L_{im}, for example, whether marijuana use is decriminalized in the market, the amount of marijuana that can be grown

[13] According to the Australian Bureau of Criminal Intelligence (1996), the increase in hydroponic systems may be related to the fact that, unlike external plantations, hydroponic cultivation is not affected by the growing seasons of the region.

[14] The baseline model is not the frequency of use rather it is the decision to use in the past 12 months.

[15] We do not include potentially endogenous covariates that may impact the utility from using cannabis such as lifetime use, education status, labor force participation, marital status, and number of children. We would need to instrument for them and the impact of these variables on cannabis use is not the primary focus of this paper. We do include health status, which may also be endogeneous to use. We run robustness checks without health status as a control variable, and the results do not change.

for a minor offense, and the (dis)taste an individual has for engaging in illegal behavior. Given that the age of an individual may influence their sensitivity to paying for marijuana or their view of doing something that is illegal ($L_{im}^{illegal}$), we also include an interaction of the age brackets (d_i^{age}) with price and with the (dis)utility of illegal behavior.[16] Specifically, the indirect utility is represented by

$$U_{im1} = p_i \alpha_1 + p_i d_i^{age\prime} \alpha_2 + d_i' \beta_1 + x_m' \beta_2 + L_{im}' \delta_1 + L_{im}^{illegal} d_i^{age\prime} \delta_2 + \epsilon_{im1}, \qquad p_i \sim \widehat{P}_m(p), \quad (1)$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2, \delta_1$ and δ_2 are (vectors of) parameters to be estimated.[17] One caveat is that we do not observe individual prices, p_i. However we know something about the distribution of the prices from the data, as discussed in section 1, which we use to construct an empirical price distribution given by $\widehat{P}_m(p)$. We discuss construction of the empirical price distribution in detail in section 3.

Individuals have utility from not using marijuana, which we model as

$$U_{im0} = x_{m0} + \epsilon_{im0}. \quad (2)$$

We normalize x_{m0} to zero, because we cannot identify relative utility levels. The $\varepsilon_{im} = \varepsilon_{im0} - \varepsilon_{im1}$ is a mean zero stochastic term distributed i.i.d. normal across markets and individuals.

One innovation of this paper is to model the role of accessibility in marijuana use.[18] We consider that whether you know where to buy it is also a function of individual i's observed characteristics and market characteristics. The probability person i has access to marijuana in market m is given by

---

[16] There may be individual characteristics that are not observed by the econometrician that impact the utility one obtains from cannabis use. We estimated specifications that include random coefficients on legality and prices. However, once we include demographic interactions there is not enough additional variation to identify the random coefficients.

[17] Our data are not longitudinal so we cannot control for (endogenous) lagged use. Therefore, one should consider our model as capturing use among recreational users and not accounting for the role possibly played by addiction. We think addiction is less of an issue for our data because, as discussed previously, our data capture mostly recreational use: only 3% report daily use or that use is a habit. However, we conduct robustness checks where we consider only non-habitual and non-daily users. These results are discussed in section 6.

[18] Our goal is to account for potential selection in use that could arise from individuals having access to cannabis. This could come from individuals searching for cannabis or being offered cannabis. We wish to get accurate estimates after controlling for selection not to understand search decisions. For this reason, as well as data limitations, we do not estimate a search model. See Galenianos, Pacula, and Persico (2009) for a theoretic search model applied to illicit markets.

$$\phi_{im} = \Pr(h_i'\gamma_1 + w_m'\gamma_2 + \eta_{im} > 0), \tag{3}$$

where h_i represents individual attributes such as whether the individual lives in a city, gender, age splines, and education variables. The market-specific variables that influence access (w_m) include arrests-per-capita for marijuana use (as a proxy of prevalence), year fixed effects, and state fixed effects. The η_{im} is an individual-market specific error term and γ_1 and γ_2 are (vectors of) parameters to be estimated.

It is likely that access to marijuana and the use decision are correlated (due to selection). For example, some individuals may have high levels of utility for using marijuana, and therefore will search for where to purchase it. This can be captured by correlation in observables (such as demographics) and correlation in the error terms in the indirect utility and access equations.

The probability that individual i chooses to use marijuana in market m depends upon the probability they know where to purchase marijuana (ϕ_{im}) and the probability they would use it given availability. Let

$$R_{im} \equiv \{U_{im1}(p_i, d_i, x_m, L_{im}, \epsilon_{im1}) \geq U_{im0}(p_i, d_i, x_m, L_{im}, \epsilon_{im0}), \phi^*_{im}(h_i, w_{m,}, \eta_{im}) > 0\}$$

define the set of variables that results in consumption of marijuana given the parameters of the model, where $\phi^*_{im} = h_i'\gamma_1 + w_m'\gamma_2 + \eta_{im}$. The probability i chooses to use marijuana in market m (the individual market share) is given by

$$\begin{aligned} S_{im} &= \int_{R_{im}} dF_{\epsilon,\eta,p}(\epsilon, \eta, p) && (4) \\ &= \int_{R_{im}} dF_{\epsilon,\eta}(\epsilon, \eta) d\widehat{P}_m(p) && (5) \end{aligned}$$

where $F(\cdot)$ denotes distribution functions, the latter equality follows from independence assumptions, and $\widehat{P}_m(p)$ represents the empirical distribution of price.

Our approach differs from the rest of the literature in a couple of fundamental ways. First, we model accessibility directly. An implicit assumption in economic models that have been considered to date is that all individuals have access to marijuana. In our framework, this is equivalent to assuming $\phi_{im} = 1$ and that there is no correlation in the observables or the errors in the indirect utility and access equations. Second, we model the (dis)utility from engaging in illegal behavior directly. We are able to do both of these things because we have data on whether individuals have access to the drug and their feelings about engaging in illegal behavior. Modeling accessibility is particularly important for drawing correct inferences about choices that individuals would make under a policy of legalization, where the accessibility issue would essentially disappear. Third, we directly address an issue that is prevalent in studies of illicit markets: the fact that prices are not observed for each purchase. To do so we use extra individual-level data on the type of marijuana used (i.e., head, hydro) combined with market-level pricing data to obtain an implied price faced by users and non-users. This allows us to estimate a model with individual prices while not observing these in the data directly.

The approach described so far informs us about the extensive margin, i.e., how people move from no marijuana use to marijuana use. We also want to get an estimate of the tax revenue that would be raised under legalization, which requires information about per unit use. Ideally, we would have information about quantity used to estimate this model. Unfortunately these data are not available, but we have information on frequency of use as discussed in section 1. We use these data to model use frequency for individual i in market m in terms of three frequencies - no use, infrequent use and frequent use. We also have information on the average amount used per session that we use to construct a quantity variable associated with each frequency. We compute tax revenues based on this information and the estimates that arise from the model of frequency of use and access. Details on how we compute tax revenues are provided in section 5.3.

3 Econometric Specification

We propose and estimate an econometric model for marijuana access and utility based on the economic model specified in the previous section. Suppose we have a sample of $i = 1, .., n$ individuals. Let $a_{im} = 0, 1$ denote whether an individual has access to marijuana ($a_{im} = 1$) or not ($a_{im} = 0$), where access to marijuana will depend on some random shock η_{im} and a vector of covariates of individual attributes and market characteristics. Here we assume that an individual's indicator of having access to marijuana can be modeled in terms of a probit

$$a_{im} = I[\mu^a_{im} + \eta_{im} > 0] \text{ where } \eta_{im} \sim N(0, 1), \tag{6}$$

where $\mu_{im}^a \equiv h_i'\gamma_1 + w_m'\gamma_2$ so that $\phi_{im} = \Pr(a_{im} = 1) = \Phi(\mu_{im}^a)$. Further, we let $u_{im} = 0, 1$ denote whether individual i has a positive (indirect) utility from using marijuana relative to the outside good. For ease of exposition, we refer to u_{im} as net-utility. We have

$$u_{im} = I[U_{im1} > U_{im0}] = I[\mu_{im}^u > \varepsilon_{im}], \tag{7}$$

where $\mu_{im}^u \equiv p_i\alpha_1 + p_i d_i^{age\prime}\alpha_2 + d_i'\beta_1 + x_m'\beta_2 + L_{im}'\delta_1 + L_{im}^{illegal} d_i^{age\prime}\delta_2$ and $\varepsilon_{im} \equiv \epsilon_{im0} - \epsilon_{im1}$.

To account for the correlation between marijuana access and use decisions as a result of unobserved confounders we assume a joint normal distribution for the two error terms and let

$$\begin{pmatrix} \eta_{im} \\ \varepsilon_{im} \end{pmatrix} \sim N\left(0, \Xi = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{8}$$

where the off-diagonal element ρ reflects the correlation between the two decisions and the diagonal elements are 1 due to the standard identification restriction for binary response variables.

In our setting with limited access, the net-utility from marijuana use is not observed for all individuals, but only reflected in the observed consumption decisions of those individuals with access. Let indicator $c_{im} = 0, 1$ denote whether consumer i is observed using marijuana. Observed consumption can be expressed in terms of access and preferences (net-utility) based on our joint model as

$$\Pr(c_{im} = 1) = \Pr(a_{im} = 1)\Pr(u_{im} = 1 | a_{im} = 1)$$

$$\Pr(c_{im} = 0) = \Pr(a_{im} = 0) + \Pr(a_{im} = 1)(\Pr(u_{im} = 0 | a_{im} = 1),$$

where $\Pr(u_{im} = j | a_{im} = 1)$ for $j = 0, 1$ is the net-utility conditional on access. The first line states that marijuana consumption reflects access to marijuana and a positive net-utility from use, while the second line shows that zero consumption could arise from: (1) no access or (2) access and negative net-utility. In other words, the observed zero consumption is inflated with zeros reflecting access only. Observing access for each individual allows us to contribute those zeros correctly to the access model. Only for individuals with access the decision whether to use marijuana reflects the net-utility from use so that for those subjects $u_{im} = c_{im}$.

Thus, we observe three possible cases, $(a_{im} = 1, u_{im} = 1)$, $(a_{im} = 1, u_{im} = 0)$ and $(a_{im} = 0)$ and the likelihood contribution for the observed access and net-utility of individual i in market m can therefore be expressed as

$$\begin{aligned}
\Pr(a_{im} = 0|\boldsymbol{\theta}) &= \Pr(\mu^a_{im} + \eta_{im} < 0) && \text{if } a_{im} = 0 \\
\Pr(a_{im} = 1, u_{im} = 0|\boldsymbol{\theta}) &= Pr(\mu^a_{im} + \eta_{im} > 0, \mu^u_{im} + \varepsilon_{im} < 0) && \text{if } a_{im} = 1, u_{im} = 0 \\
\Pr(a_{im} = 1, u_{im} = 1|\boldsymbol{\theta}) &= Pr(\mu^a_{im} + \eta_{im} > 0, \mu^u_{im} + \varepsilon_{im} > 0) && \text{if } a_{im} = 1, u_{im} = 1
\end{aligned}$$

where $\boldsymbol{\theta}$ refers to the vector of all model parameters. In other words, given our normal error specifications we have a univariate probit for access for subjects with no access and a bivariate probit for access and net-use for subjects with access so that we can rewrite the above expressions for the likelihood contribution of individual i more compactly as

$$\Phi(-\mu^a_{im})^{(1-a_{im})} \; + \; \Phi_2(\mu^a_{im}, -\mu^u_{im}; \rho)^{(a_{im})(1-u_{im})} + \; \Phi_2(\mu^a_{im}, \mu^u_{im}; \rho)^{(a_{im})(u_{im})} \tag{9}$$

where $\Phi(\cdot)$ refers to the CDF of a standard normal distribution and $\Phi_2(\cdot)$ to the CDF of a standard bivariate normal distribution.

Based on these expressions we can identify the parameters for the access and the net-utility models and the correlation. The exclusion restrictions are the arrests-per-capita by state (which are a proxy for the prevalence of marijuana)[19] and whether the consumer lives in a major city, both of which may impact accessibility but are assumed not to impact utility; and the health status of the individual, whether the marijuana is of high quality, and the (dis)taste of doing something that is illegal, which may impact utility but not accessibility.

As mentioned previously, we do not observe individual prices, p_i. We use data on the distribution of market prices by marijuana type, $\mathbf{p}_m = \{p^{leaf}_m, p^{head}_m, p^{hydro}_m\}$, and a vector of use by type from the information on the type of marijuana each individual uses, $\boldsymbol{\pi}_m = \{\pi^{leaf}_m, \pi^{head}_m, \pi^{hydro}_m\}$. We use these data to generate an empirical distribution of price across individuals that varies by market given by

$$\widehat{P}_m(p) = \int g(\mathbf{p}_m, \boldsymbol{\pi}_m) \; dF_{(\boldsymbol{\pi}_m)} dF_{(\boldsymbol{p}_m)}, \tag{10}$$

where $g(\cdot)$ is a function that creates an average price that an individual would have faced over the past year based on his or her type of usage and type prices, which are coming from

[19] Arrests-per-capita refer to arrests of suppliers, not users. For this reason, arrests-per-capita are unlikely to impact the utility associated with using cannabis but are likely to impact the prevalence of cannabis for sale.

distributions $F_{(\boldsymbol{\pi}_m)}$, and $F_{(\boldsymbol{p}_m)}$ (centered at the corresponding $\boldsymbol{\pi}_m$ and $\mathbf{p}_m$), respectively.[20] This approach improves upon the use of average market prices, enabling us to obtain the implied price faced by users and non-users in a symmetric way and to properly address the econometric issue of unobserved individual prices in estimation by integration.

As the prices are not individual reported purchase price there may be some concern that price is correlated with the error term, and, therefore endogenous. As discussed in Section 1, prices are higher the higher is potency, which can be thought of as measure of the quality of the marijuana. We include a measure of the potency to control for quality to ameliorate the concern that prices are correlated with the error term (when the error term captures unobserved quality). We also conduct robustness checks to further investigate price endogeneity after controlling for quality. These details can be found in section 6.

Let $\mathbf{a}_m = \{a_{1m}, ..., a_{n_m m}\}$ denote the vector of access variables for all n_m subjects in market m, $\mathbf{u}_m = \{u_{1m}, ..., u_{n_{1_m} m}\}$ the vector of net-utility variables for the n_{1_m} subjects in market m with access to marijuana and $\mathbf{W}_m = \{\mathbf{W}_{1m}, ..., \mathbf{W}_{n_m m}\}$ the matrix of all covariates excluding price. Grouping subjects in each market by marijuana access, we define the sets I_{m1} for all subjects with access and I_{m0} for all subjects with no access. The likelihood of observing the data $(\mathbf{a}_m, \mathbf{u}_m)$ for all subjects in market m can then be be expressed in two parts for the set of non-access and access subjects as

$$f(\mathbf{a}_m, \mathbf{u}_m|\boldsymbol{\theta}, \mathbf{W}) = \prod_{I_{m0}} \Pr(a_{im} = 0|W_{im}, \boldsymbol{\theta}) \prod_{I_{m1}} \int \Pr(a_{im} = 1, u_{im} = j|W_{im}, \boldsymbol{\theta}, p_i)\, d\widehat{P}_m(p), \tag{11}$$

where p_i is the individual-specific price coming from the distribution defined in equation (10). The expression under the integral is the term $\Phi_2(\mu^a_{im}, -\widetilde{\mu}^u_{im}; \rho)^{(a_{im})(1-u_{im})} + \Phi_2(\mu^a_{im}, \widetilde{\mu}^u_{im}; \rho)^{(a_{im})(u_{im})}$ from (9) where the mean term $\widetilde{\mu}^u_{im}$ uses the price p_i. For all individuals in the sample the likelihood is simply a product over the likelihoods for all markets $m = 1, ..., M$, $f(\mathbf{a}, \mathbf{u}|\boldsymbol{\theta}, \mathbf{W}) = \prod_{m=1}^{M} f(\mathbf{a}_m, \mathbf{u}_m|\boldsymbol{\theta}, \mathbf{W})$, where $\mathbf{a} = \{a_1, ..., a_M\}$, $\mathbf{u} = \{u_1, ..., u_M\}$ and $\mathbf{W} = \{\mathbf{W}_1, ..., \mathbf{W}_M\}$ refer to the observed data for all sample subjects.

We also consider an extended version of the above model for the ordered marijuana use response variable $y_{im} = k$, $k = 0, 1, 2$ for all individuals with access, where the three categories refer to "no use", "infrequent use" and "frequent use," respectively. Extending

[20] We assume that the distributions of prices and market usage are independent across types. Alternatively, we could allow for correlation in prices across types, across usage of types, and/or correlation in the joint distribution of prices and type, however this would require augmenting the parameter space to include at least 3 more covariance parameters for each of the 27 markets.

the model for use given in equation (7) above, we have

$$y_{im} = 0 \text{ if } (\mu^u_{im} + \nu_{im}) \leq 0,\ y_{im} = 1 \text{ if }\ 0 < (\mu^u_{im} + \nu_{im}) \leq \tau \text{ and } y_{im} = 2 \text{ if }\ \tau < (\mu^u_{im} + \nu_{im})$$

where τ is a cut-off parameter to be estimated and ν_{im} refers to the random shock in the latent utility of marijuana use in the ordered probit model. The mean μ^u_{im} depends as before on a set individual characteristics such as demographics and price, market specific variables, and legality related variables. As in the bivariate probit model above we allow for selection based on unobservables and model the access and use decision jointly. We again assume a joint normal distribution of the error terms of the use and access model, $(\eta_{im}, \nu_{im}) \sim N(0, \Xi)$, to allow for the correlation in unobservables, with the access model specified as before (see equation 6). Under the ordered probit outcome for marijuana the likelihood contribution of individual i we now have

$$\Pr(a_{im} = 1, y_{im} = 1|\boldsymbol{\theta}) = Pr(\mu^a_{im} + \eta_{im} > 0, \mu^u_{im} + \nu_{im} < 0) \qquad \text{if } a_{im} = 1, y_{im} = 0$$

$$\Pr(a_{im} = 1, y_{im} = 1|\boldsymbol{\theta}) = Pr(\mu^a_{im} + \eta_{im} > 0, 0 < \mu^u_{im} + \nu_{im} < \tau) \qquad \text{if } a_{im} = 1, y_{im} = 1$$

$$\Pr(a_{im} = 1, y_{im} = 2|\boldsymbol{\theta}) = Pr(\mu^a_{im} + \eta_{im} > 0, \tau < \mu^u_{im} + \nu_{im}) \qquad \text{if } a_{im} = 1, u_{im} = 2$$

As before we have $\Pr(a_{im} = 0|\boldsymbol{\theta}) = \Pr(\mu^a_{im} + \eta_{im} < 0)$ for non-access subjects. Addressing the issue of the unobserved individual prices as described above, the likelihood contribution of all subjects in market m is

$$f(\mathbf{a}_m, \mathbf{y}_m|\boldsymbol{\theta}, \mathbf{W}) = \prod_{I_{m0}} \Pr(a_{im} = 0|W_{im}, \boldsymbol{\theta}) \prod_{I_{m1}} \int \Pr(a_{im} = 1, y_{im} = k|W_{im}, \boldsymbol{\theta}, p_i)\, d\widehat{P}_m(p)$$

where $\mathbf{y}_m = \{y_{1m}, ..., y_{n1_m m}\}$ is the vector of the ordered response variable on frequency of use for all subjects in market m.

We estimate both models via standard Bayesian Markov Chain Monte Carlo methods, building closely on Chib and Jacobi (2008) and Bretteville-Jensen and Jacobi (2011). The details are provided in Appendix A. Bayesian methods are increasingly used in empirical analysis including in empirical IO (see for example Jiang, Manchanda and Rossi, 2009).[21] The methods are well suited to deal with discrete response variables and the more complex likelihood structure arising from the joint modeling of marijuana use and access. Specific

[21] We also estimated the baseline models using frequentist MLE methods. We obtain the same results for the parameter point estimates (mean) and standard error (standard deviation) up to three decimal places of precision.

to our context, the Bayesian approach enables us to address the issue of dealing with unobserved individual prices in a realistic and flexible approach described above and, in addition, provides a natural framework to implement our counterfactual analysis of marijuana use under legalization.

4 Results

In this section we discuss the role that access plays in marijuana use and the importance of correcting for selection into use. We also examine age-related differences in sensitivity to policy variables such as price and legality. In all specifications we use the intermediate definition of the level of access (i.e., access 2 in Table 1). We present robustness checks using two other definitions of access in Section 6.

Table 4 presents results from two probit models of marijuana use in the first two columns and results from a model corrected for selection in the last two columns. As we discussed earlier, previous literature has not accounted for selection, therefore we refer to the results from the probit models as coming from the standard approach. The first probit model includes a dummy variable indicating whether marijuana use is decriminalized in the market. The other specifications include state fixed effects. Both probit models indicate that males and individuals in their teens and twenties are more likely to use marijuana relative to females and other age categories and that use is declining with age. They also indicate that aboriginal individuals are more likely to use, while those who report being in better health are less likely to use. For both probits, estimates of individual attribute parameters are similar. However, the estimates vary with respect to market variables, which could be attributed to differences across states that are not controlled for in the decriminalization specification (for example, variation in enforcement of marijuana laws). Therefore, we focus on models that include state fixed effects in the remainder of the analysis.

Estimates from the bivariate model with selection correction (with state fixed effects) are presented in Columns 3 and 4. The results illustrate that access is not randomly distributed across individuals, the same observables impact access and use conditional on access. Specifically, age has a different magnitude of impact for access and use. In addition, there is selection on unobservables. This is reflected by the fact that the distribution of the correlation in unobservables (ρ) is positive. It is centered at 0.2 with more than 90% of the distribution in the positive range. These results indicate that it is important to correct for selection in marijuana use because accessibility is non-random across individuals.

	Probit		Bivariate Probit with Selection	
	(1)	(2)	(3)	(4)
	Decriminalized	State Fixed Effects	State Fixed Effects	
	Cannabis Use	Cannabis Use	Cannabis Use	Access
Individual Attributes				
Male	0.330 *	0.332 *	0.307 *	0.291 *
	(0.015)	(0.015)	(0.022)	(0.012)
Age in Teens Spline	0.153 *	0.154 *	0.119 *	0.140 *
	(0.017)	(0.017)	(0.019)	(0.016)
Age in Twenties Spline	-0.033 *	-0.034 *	-0.029 *	-0.039 *
	(0.003)	(0.003)	(0.004)	(0.003)
Age in Thirties Spline	-0.039 *	-0.039 *	-0.031 *	-0.040 *
	(0.003)	(0.003)	(0.004)	(0.003)
Age in Forties Spline	-0.029 *	-0.029 *	-0.026 *	-0.020 *
	(0.003)	(0.003)	(0.004)	(0.002)
Age over Forties Spline	-0.078 *	-0.078 *	-0.073 *	-0.054 *
	(0.005)	(0.005)	(0.007)	(0.003)
Highest Education is High School	-0.100 *	-0.091 *		-0.006
	(0.024)	(0.025)		(0.020)
Highest Education is Trade Degree	0.001	0.004		0.110 *
	(0.020)	(0.020)		(0.016)
Highest Education is University Degree	-0.128 *	-0.117 *		-0.019
	(0.023)	(0.023)		(0.017)
Of Aboriginal Descent	0.199 *	0.163 *	0.157 *	0.211 *
	(0.050)	(0.051)	(0.057)	(0.047)
In Good, Very Good, or Excellent Health	-0.282 *	-0.285 *	-0.267 *	
	(0.015)	(0.015)	(0.017)	
Market and Policy Variables				
Price	-0.001	-0.003	-0.004 *	
	(0.001)	(0.001)	(0.002)	
Use if Legal	0.467 *	0.467 *	0.379 *	
	(0.024)	(0.024)	(0.026)	
Grams Possession is not Minor Offense	-0.001 *	0.000	0.001	
	(0.000)	(0.001)	(0.000)	
Arrests Per Capita of Suppliers (Prevalence)	0.213 *	0.239 *		-0.008
	(0.048)	(0.084)		(0.068)
Live in City	0.000	-0.023		-0.187 *
	(0.016)	(0.016)		(0.013)
High Potency	-0.022	-0.176	-0.0418	
	(0.144)	(0.149)	(0.163)	
Decriminalized	0.151 *			
	(0.020)			
Correlation (ρ)			0.223	
			(0.106)	
Number of Observations	51248	51248	51296	51296

Notes: Standard deviations in parentheses. * indicates 95% Bayesian confidence interval does not contain zero.
All specifications include year fixed effects and a constant in access and use. State fixed effects are included
in access and use equations for the state fixed effects spec.

Table 4: Estimates of Selection Models and Probits

Selection results indicate that individuals that live in a city are less likely to have access to marijuana. This is consistent with the reported growing patterns of marijuana in Australia, where it is usually grown in sparsely populated areas ("the outback") and hence it is easier to obtain outside of cities. Access results indicate that, conditional on age, individuals whose highest education is a trade degree are more likely to have access. Given that we control for state fixed effects, we use the supplier arrest rate as a proxy for the prevalence of marijuana in the market (as police enforcement is usually consistent within states). The results indicate higher prevalence is consistent with higher access.

The results from the selection model also yield different interpretations of the impact of policy variables on use as well as different patterns of behavior with respect to age groups. Specifically, the probit and the selection model differ in that the former suggests individuals are less sensitive to price and more sensitive to legalization laws than a model corrected for selection indicates. This is of particular importance as these are market specific variables that the government can control through policy and hence impact the probability of use. The selection model indicates participation is more elastic with respect to price: a 10% increase in price results in a 2.2% decrease in the probability of use while the standard model indicates the probability of use would drop by a lower amount (1.7%). The magnitude of the participation elasticity (-0.22) from the selection model is consistent with estimates of cigarette participation elasticities from prior studies (which range from -0.25 to -0.50).[22] This similarity is not surprising as marijuana is combined with tobacco when consumed in Australia.

Table 5 presents selected parameter results of three models with age interactions (for the specification with state fixed effects in use and access). All specifications include the same control variables as in Table 5. For ease of comparison we reproduce the relevant results for the specification without age interactions in the first panel. The second panel shows estimates from price and age interactions. The results indicate that there is variation in price sensitivity across age groups. Teenagers are somewhat less sensitive to price changes than older individuals. This implies that increases in prices (via a tax, for example) will have less of an impact on use among younger individuals. The third panel indicates that there is age variation in the disutility of participating in illegal activities. Again, teenagers exhibit the least sensitivity to the legal status of marijuana of all age groups, where the sensitivity to legal status is increasing in age, with exception of the highest age group. The final panel presents results with interactions of age with price and legality. The results mirror those

[22] See the literature review in Chaloupka et al (2002) and Chaloupka and Warner (2000).

of the previous specifications. Overall, the findings indicate that variables associated with legality and prices (two policy instruments) will both have less of an impact on teenagers and individuals in their twenties relative to other age groups. The results for the ordered probit model of frequency of use (that we use to compute tax revenues) yield similar patterns across age groups. However, the elasticity estimates are higher since these are elasticities associated with frequency of use rather than participation. We present the frequency ordered probit with selection results in Appendix A.3.

	Bivariate Probit with Selection (with State Fixed Effects in Use and Access)							
Interactions:	No interactions (Table 5)		Price and Age		Legality and Age		Legality, Price and Age	
	Cannabis Use	Access	Cannabis Use	Access	Cannabis Use	Access	Cannabis Use	Access
Age Splines								
Age in Teens	0.119 *	0.140 *	0.138 *	0.141 *	0.106 *	0.141 *	0.128 *	0.140 *
	(0.019)	(0.016)	(0.024)	(0.016)	(0.020)	(0.016)	(0.025)	(0.016)
Age in Twenties	-0.029 *	-0.039 *	-0.019 *	-0.039 *	-0.030 *	-0.039 *	-0.020 *	-0.039 *
	(0.004)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)	(0.003)
Age in Thirties	-0.031 *	-0.040 *	-0.029 *	-0.040 *	-0.032 *	-0.040 *	-0.029 *	-0.040 *
	(0.004)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.006)	(0.003)
Age in Forties	-0.026 *	-0.020 *	-0.036 *	-0.020 *	-0.026 *	-0.020 *	-0.036 *	-0.020 *
	(0.004)	(0.002)	(0.006)	(0.002)	(0.004)	(0.002)	(0.006)	(0.002)
Age over Forties	-0.073 *	-0.054 *	-0.074 *	-0.054 *	-0.070 *	-0.054 *	-0.074 *	-0.054 *
	(0.007)	(0.003)	(0.009)	(0.003)	(0.007)	(0.003)	(0.009)	(0.003)
Price and Interactions:								
Price	-0.004 *				-0.004 *			
	(0.002)				(0.002)			
Age in Teens			-0.001				-0.001	
			(0.002)				(0.002)	
Age in Twenties			-0.004 *				-0.004	
			(0.002)				(0.002)	
Age in Thirties			-0.006 *				-0.006 *	
			(0.002)				(0.002)	
Age in Forties			-0.003				-0.004	
			(0.002)				(0.002)	
Age over Forties			-0.002				-0.001	
			(0.002)				(0.002)	
Use if Legal and Interactions:								
Use if Legal	0.379 *		0.382 *					
	(0.026)		(0.026)					
Age in Teens					0.185 *		0.173 *	
					(0.061)		(0.061)	
Age in Twenties					0.323 *		0.325 *	
					(0.046)		(0.046)	
Age in Thirties					0.475 *		0.492 *	
					(0.049)		(0.051)	
Age in Forties					0.571 *		0.567 *	
					(0.063)		(0.064)	
Age over Forties					0.314 *		0.305 *	
					(0.089)		(0.088)	

Notes: Standard deviations in parentheses. * indicates 95% Bayesian confidence interval does not contain zero. Includes all controls in Table 5 including individual attributes, year and state fixed effects in use and access.

Table 5: Selected Parameter Estimates for Price, Age, and Illegality Interactions

5 Policy Analysis

We use the results from the selection model to investigate the effect of legalization and to improve our understanding about individual's decision making in that context. We aim to address the following policy concerns: (i) what role does access play in marijuana use; (ii) what role do other factors (such as demographic characteristics, illegality of the drug, prices, etc.) play in the decision to use the drug, (iii) can we use policy to restrict use among young adults, and (iv) how does legalization impact tax revenues.

5.1 Impact of Accessibility and Legalization on Use

We decompose the impact of legalization in three ways: the part of the increase in use due to increased accessibility, the part due to the removal of the stigma associated with breaking the law, and the part due to potential changes in prices due to supply side cost changes or tax policies. Specifically, if marijuana were legalized, then accessibility would not be as large of a hurdle. In the model very easy access implies $\phi_{im} = 1$. In addition, the disutility associated with illegal activity would be zero; in the model this implies $L_{im}^{illegal} = 1$.[23] Furthermore, dealers would no longer face penalties for selling. To address this issue, we compute the counterfactuals under various assumptions about how price would change: (i) price would not change; (ii) price would increase by 25%; (iii) price would decline to the price of cigarettes; and (iv) price would decline to the marginal costs of production. Notice that since we don't model the supply side prices are taken as exogenous. In all scenarios, we change the environment and recompute the predicted probability of use that would arise in the counterfactual world implied by the parameter estimates from the state fixed effects baseline specification (presented in Table 4 columns 3 and 4).

We discuss our choice of counterfactual prices in turn. The first scenario (no change in price) is not realistic, however it serves as a benchmark to our other counterfactuals. Scenario (ii), a 25% price increase, is motivated by tax proposals made in the United States, where legalization laws were recently passed. Specifically, in 2013 Washington state legalized marijuana use for recreational purposes. Selling remains illegal, but the relevant amendment, I-502, gives state lawmakers until 2015 to establish a system of state-licensed growers,

[23] Notice that our model allows for selection on unobservables (as well as observables). When we estimate the model each person will have a realization of the unobserved term for each iteration (which is correlated with use according to $\widehat{\rho}$). For each individual we use the average of their unobserved terms along with observables when computing the counterfactuals.

processors and retail stores, where they propose to tax marijuana 25%. Scenario (iii) is more reasonable as marijuana is typically mixed with tobacco in Australia. Finally, the last scenario, pricing at marginal cost, serves as a lower bound on the price of marijuana. We use marijuana marginal production cost estimates reported in Caulkins (2010). These estimates are based on the costs for growing other herbs (e.g., the price of plants, growing fertilizer, labor, etc.).[24]

Environment			Predicted Probability of Use For Current Consumers: All		With No Access		With Access	
Accessible	Legal	Price	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
No Change	No	No Change	0.139		0.000		0.266	
Accessible	No	No Change	0.191	(0.135)	0.110	(0.081)	0.266	(0.130)
Accessible	Yes	No Change	0.287	(0.162)	0.186	(0.112)	0.380	(0.144)
		25% Increase	0.272	(0.158)	0.174	(0.108)	0.363	(0.142)
		Cigarette	0.349	(0.178)	0.238	(0.131)	0.450	(0.152)
		Cost	0.349	(0.178)	0.238	(0.131)	0.450	(0.152)

Notes: Based on baseline specification with state fixed effects. The first row is a prediction for a person with the typical access characteristics. All estimates fall within the 95% Bayesian Confidence (Credibility) Intervals

Table 6: Counterfactual Use Results

Table 6 displays the counterfactual results which indicate that both access and illegality concerns play substantial roles in the decision to use marijuana. The first row replicates the data under the current legal environment. The second row shows how the probability of use would change if accessibility were not an issue in an environment where use was still illegal. That is we assume all other aspects of the counterfactual world stay the same other than access, so we recompute the probability of use assuming that $\phi_{im} = 1$ for all individuals. In this scenario, the probability of use among current non-users without access would increase to 11% resulting in an overall increase of 37% in probability of use (from 13.9% to 19.1%). If marijuana were legalized (i.e., we set $L_{im}^{illegal} = 1$) and accessibility were not an issue use would more than double to 28.7%. Obviously there would be an impact on prices due to

[24] Caulkins reports that, in the US, wholesale prices range from \$500 to \$1500 per pound. Due to electrical usage costs of growing hydro are higher, between \$2000–4500 per pound. If cannabis is grown outside production costs are estimated to be less than \$20 per pound. The costs in Australia are likely to be of the same magnitude as the costs of low-skilled labor and raw inputs are similar to those in the US.

the law change, however, even if prices increased by 25% the probability of use would still rise to almost double current levels.

5.2 Legalization and Use among Young Adults

Finding ways to limit use of drugs among young adults is an important issue in the legalization debate. As the results from Table 5 show, the impact of prices and legality varies by age group. We use these estimates to compute counterfactual use probabilities by age group. This allows us to conduct various age-specific counterfactuals that can give us insight about the prevalence of use among youths in a legalized setting.

Environment			Predicted Probability of Use For Individuals in Age Bracket:					
Accessible	Legal	Price	All	Teen	Twenties	Thirties	Forties	Over Forty
No Change	No	No Change	0.139	0.274	0.264	0.151	0.102	0.038
Accessible	No	No Change	0.191	0.347	0.329	0.217	0.169	0.077
			(0.135)	(0.130)	(0.118)	(0.098)	(0.086)	(0.057)
Accessible	Yes	No Change	0.287	0.464	0.457	0.33	0.27	0.139
			(0.162)	(0.128)	(0.120)	(0.113)	(0.106)	(0.082)
		25% Increase	0.272	0.46	0.441	0.308	0.258	0.135
			(0.158)	(0.129)	(0.120)	(0.110)	(0.104)	(0.081)
		Cigarette	0.349	0.482	0.52	0.419	0.319	0.157
			(0.178)	(0.132)	(0.123)	(0.124)	(0.117)	(0.091)
Smoked Cigarette in the Past Year			48.6%	84.4%	71.3%	51.6%	43.2%	37.2%
Daily Cigarette Smoker			29.2%	25.7%	35.1%	31.2%	29.4%	26.1%
Report Current Access to Cannabis			53.0%	71.5%	73.3%	58.5%	48.1%	32.8%

Notes: This is a prediction of use for a person with the typical access characteristics using estimates from the state fixed effects specification with price-age interactions. Standard deviations are in parethesis. All estimates fall within the 95% Bayesian confidence (credibility) intervals.

Table 7: Counterfactual Use Results by Age Group

Table 7 presents the counterfactual results by age group. The results indicate that if marijuana were legal and priced the same as cigarettes the probability of use among teens would increase by three-fourths (from 27.4% to 48.2%) and would nearly double among twenty year olds (from 26.4% to 52%). However, the probability of using marijuana in a year would not be as high as the probability of smoking a cigarette in a year (average of 34.9% relative to 48.6%). The results also indicate that legalizing marijuana would have a smaller relative impact on younger individuals. If a 25% tax over current prices was implemented then the probability of use for individuals in their teens and twenties would increase by around 67% post legalization, while the probability would more than double in other age groups.

The results also indicate that the impact of accessibility on use probability relative to legality differs by age group. Those in their teens and twenties exhibit almost the same pre-legal levels of access to marijuana (71.5% and 73.3%, respectively). Likewise, they react similarly to the removal of the access barrier, which sees growth in the probability of use of around 23% for teens (from 27.4% to 34.7%) and 24% for those in their twenties (from 26.4% to 32.9%). In contrast, the impact of legalization alone increases use by an additional 33% for teens (from 34.7% to 46.4%), while other age groups are impacted more. For example, among individuals in their twenties the probability of use increases by 39% (rising from 32.9% to 45.7%) and individuals in their thirties have an increased probability of use of more than 50%. These results suggest that non-teens are impacted more while marijuana use is illegal, and, hence, once this barrier is removed they would increase their use relatively more than teenagers.

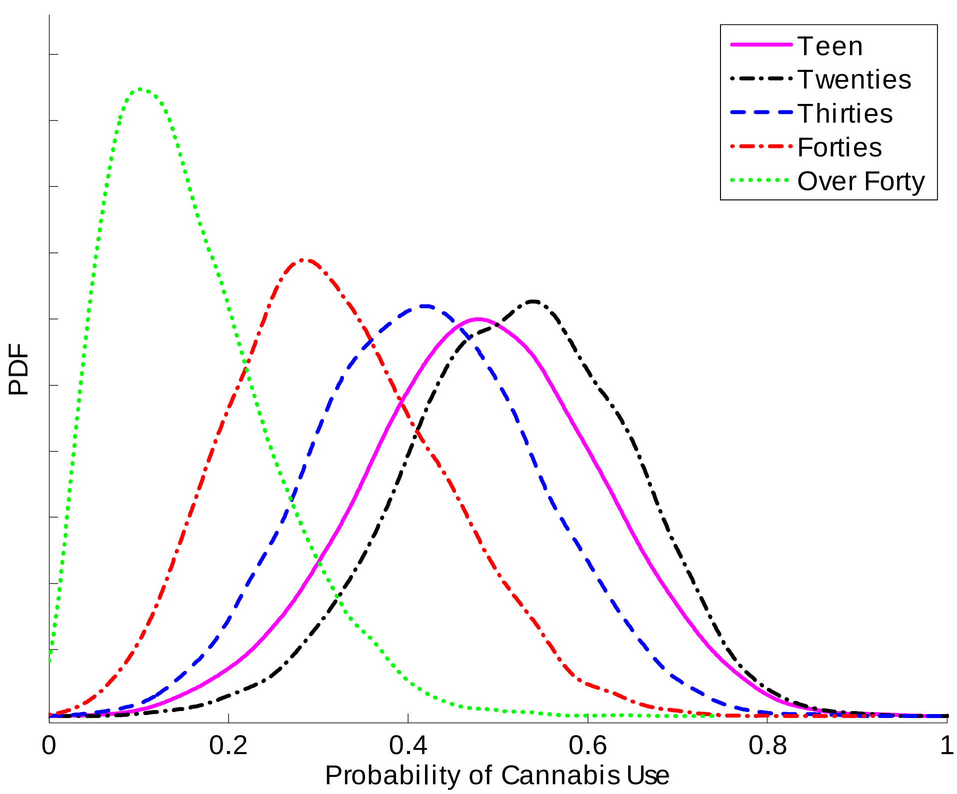

Figure 1: Predicted Use Probablity Distributions (Legalized; 25% Price Increase)

In addition to variation across the mean in use, the shape of the probability of use distribution varies by age across all legalization scenarios. For example, Figure 1 presents the distribution of the probability of use under the counterfactual of legalized marijuana with 25% higher price. As the figure illustrates the age distributions are centered at different means but also have different shapes. Specifically, we see that the distribution of use among individuals below their forties are more dispersed than the distributions for individuals over the age of forty (green dashed line). It is important to consider both the mean as well as the dispersion in use among young individuals to get an accurate picture of how use would change post-legalization.

The 67% post-legalization increase in the probability of teenage use can be viewed as an increase upper-bound as it would be the result if marijuana were legal and freely accessible to all teenagers. However it is likely that underage users would face the same (if not more harsh) restrictions on marijuana use as they face for alcohol use, which is illegal for those aged under 18. In other words, obtaining marijuana would continue to be a hurdle for underage users in the same sense that obtaining alcohol or cigarettes currently is. To accurately depict the most likely post-legalization environment, we conduct two further counterfactuals focused on individuals under the age of 18. In the first counterfactual we compute the probability of underage use assuming it is illegal but still freely accessible for the underaged; the second counterfactual adds the extra hurdle that access is also restricted. In both counterfactuals we assume marijuana is priced at cigarette prices (so a lower price). In other words, the second counterfactual is the same as the first row of Table 7 for underage users (i.e., 16 and 17 year olds) but now users face lower prices. We view the second counterfactual as a lower-bound for underage use post-legalization. In the current pre-legal environment, the probability of underage use is 23.1%. We find that the probability of underage use would increase by 38% from pre-legalization levels (to 31.8% from 23.1%) if it were made freely accessible and would increase by 5% (to 24.4%) under the second counterfactual scenario (i.e., restricted access with lower prices).

One of the policy variables that the government can use is the amount of taxes to impose. We conducted a policy experiment to determine if the government could restrict use among teenagers to pre-legalization levels by imposing a (reasonable) tax on marijuana. That is, we computed the distribution of prices that would be necessary such that the distribution of post-legalization probability of teenage use is as close as possible to the pre-legalization distribution of probability of teenage use. Details on how we compute the counterfactual price distributions are given in Appendix B. We found that prices would need to be more than ten-fold higher than current levels in order to keep the frequency of post-legalization use the same as pre-legalization use even if underage users would still face the same restrictions as they face for alcohol use. Increasing prices by ten-fold is not feasible given that we would expect most users to resort to the black market. Hence, our results indicate that, while teens respond to prices, an environment where use is legalized will see an increase in the probability of use of at least 5% and at most 67% among teens.

5.3 Tax Revenues and the Black Market

We use the (selection corrected) ordered probit model of use frequency to compute annual tax revenue under two taxation schemes. The first taxation scenario involves imposing the cigarette tax rate assuming the base price is marginal cost. In this case the consumer would pay Tax_1 over marginal cost, where marginal cost depends on the quality of marijuana purchased. The second scheme is motivated by the proposal in the US to tax marijuana at 25% over current prices. In this case $Tax_2 = 0.25$, where the price paid by consumers is $1.25p_i$. Notice the first scenario involves a lower price than currently paid and the second scenario a higher price. Together these two tax scenarios should provide a reasonable idea of the bounds on tax revenue that could be generated.

We compute annual tax revenue in three steps. First, we use the parameters of a model of frequency of use to compute the utility associated with each level of use if marijuana were legal (and hence easily accessible) under the two prices implied by the tax scenarios. Second, we compute the probability that an individual's consumption falls into one of the three frequency categories. Let $\widehat{F}_{ij}$ be a 3-dimensional vector where element j is the probability individual i's predicted frequency falls in category j. Finally, we compute the tax revenue that would be realized under the counterfactual prices according to

$$\text{Tax Revenue}_j = \begin{cases} 0 & \text{when } j = 0 \\ Tax_1 * \widehat{F}_{i1} * [1,4] * q_{im} & \text{when } j = 1 \\ Tax_2 * \widehat{F}_{i2} * [12,365] * q_{im} & \text{when } j = 2 \end{cases}, \qquad (12)$$

where Tax_j is the per-unit (gram) tax charged as described above.[25] The other terms represent total quantity consumed for each type of frequency use.

Total quantity consumed depends on the frequency of consumption, given by $\widehat{F}_{ij}$. An infrequent user ($\widehat{F}_{i1}$) is one who uses once quarterly, biannually, or annually. A frequent user ($\widehat{F}_{i2}$) is one who uses monthly, weekly, or daily. The intervals represent the lower and upper bounds on the units of consumption associated with the frequency definition (e.g., the interval takes on the lower value of 12 for a frequent (monthly) consumer who consumes 12 times per year). We compute upper and lower bounds on tax revenues as implied by these intervals.

The q_{im} denotes the grams of marijuana consumed on average per session. This is available in the data on an individual level for those individuals that consume so we can construct an average per market (q_m). We compute an average quantity consumed for all individuals (even those who didn't consume pre-legalization) as follows

$$q_{im} = \int q_m \, dF_{(\boldsymbol{q}_m)}, \tag{13}$$

where the truncated distribution $F_{(\boldsymbol{q}_m)}$ is centered at the mean of the data q_m.[26] Notice that the implicit assumption is that the average amount consumed per using session does not change. That is, we assume price changes influence quantity through use frequency but not the amount consumed per session. For example, perhaps a user smokes a marijuana joint during a party once per month. We assume that the price may change the frequency with which the user smokes (to once every few months for example) but when he smokes he still consumes one joint.

For a population the size of the US, our results indicate that a policy of marijuana legalization would raise a minimum amount between \$132 million to \$2.5 billion annually, depending on which taxation scheme is used and assuming individuals consume at the lower level of the frequency interval. It is less likely all individuals consume at the upper end of

[25] The frequency model predicts a certain probability a person lies in each of the 3 intervals. For example, a hypothetical person 1 falls into category $j = 0$ 10% of the time, into $j = 1$ 70% of the time, and into $j = 2$ 10% of the time. So the quantity consumed for this person is 0, 10% of the time; $[1, 4] * q_{im}$, 70% of the time; and $[12, 365] * q_{im}$, 10% of the time.

[26] Note that, since we have frequency of use in brackets for each person, we compute three separate quantity consumed per person: one computed at lower bound of frequency (so in this example 1 and 12), one at mean of interval; and one at the upper bound of the interval (4 and 365). We then can compute an average tax per user. To compute the average tax per user we consider a person a user if the predicted frequency they fall into category $j = 0$ is lower than 50%.

the frequency interval (i.e., to do so would mean all monthly users are treated as daily users), but if individuals consume at the midpoint of the frequency intervals then tax revenues would increase to between \$557 million to \$10 billion annually depending on the taxing scheme. We should note that our findings are consistent with those from a 2005 report funded by the Marijuana Policy Project (Miron, 2005) which estimates legalization would raise tax revenues of \$2.4 billion annually if it were taxed like most consumer goods and over \$6 billion annually if it were taxed similarly to alcohol or tobacco.

According to the Framework Convention on Tobacco Control (2012), illicit trade in cigarettes accounts for approximately one-tenth of global sales. Likewise, it is reasonable to conjecture that some marijuana users will purchase from the black market, especially if tax rates are high. We compute an adjusted tax revenue that allows for some sales to be lost to the black market. If we assume that 10% of the sales will be lost to the black market, tax revenues would decline to between \$502 million to \$9.2 billion (at the midpoint of the frequency). We also computed how much tax revenue would be raised if all users who currently use (i.e., those who are currently willing to do something illegal) would buy on the black market instead of in the legal market. In this situation tax revenues would be between \$452 million and \$8 billion (at the midpoint of the frequency).

To summarize, in the worst case tax revenue scenario - all current users purchase on the black market - legalization would still result in tax revenues of half a billion annually. At the other extreme, the government would raise \$10 billion in taxes. Furthermore, governments would see cost reductions under legalization as they would not incur nearly as high of costs of enforcement.

6 Robustness Checks

We conducted a number of robustness checks of our results. First, given the importance of the role played by access in our results we reran our baseline and interaction specifications using two different definitions of the accessibility variable. The first access variable definition is more inclusive (access 1 in Table 1) the second is more restrictive (access 3 in Table 1). The results using either access definition are virtually identical to those in the main text of the paper - they indicate that there is selection on observables and the parameter estimates are almost identical for all variables. The only notable difference is the estimate of the ρ parameter. Both alternative access definitions yield a more positive ρ : it is centered at

0.334 (0.391) with a standard deviation of 0.12 (0.83) in the access 1 (access 3) definition relative to the access 2 definition (centered at 0.223 with a standard deviation of 0.10). These results indicate that the definition of access is not driving our results - indeed the intermediate level of access produces the least selection problem relative to a more lenient or a more restrictive definition.[27]

We also conducted two model specification checks. In the first we consider the role of addiction or habitual use on current consumption. As discussed previously, our data are not longitudinal and hence we do not have information on use in previous periods so we cannot include a lagged (endogenous) use variable in the regression. However, we do observe in the data the frequency with which individuals use marijuana. Approximately 3% of the sample report using daily or that use is a habit. We reran the regressions excluding these individuals. The results are the same as those we obtain when we include this group, with one notable change: the variable "aboriginal" is no longer significant. These results suggest that our results are not being driven by the impact of habitual users. In the second specification check we consider that there may be individual characteristics that are not observed by the econometrician that impact the utility one obtains from marijuana use. We estimated specifications that include random coefficients on legality and prices. However, once we include demographic interactions there is not enough additional variation to identify the random coefficients.

Finally, we ran robustness checks of our results to the potential endogeneity of some covariates. The first concerns the endogeneity of health status where a potential concern is reverse causality - use influences health status. We reran our baseline specification without health status and there are no notable changes in the results. Second, as the prices are not individual reported purchase price there may be some concern that price is correlated with the error term, and, therefore endogenous. We include a measure of the potency to control for quality to ameliorate this concern. As discussed in Section 1, prices are higher the higher is potency, which can be thought of as measure of the quality of the marijuana. In our setting for price endogeneity to be an issue it would be necessary for something unobserved (and hence in the error term) that impacts pricing decisions and that also matters to the consumer that is not related to quality. Furthermore, if we had access to individual price paid then there would have to be something that impacted marginal costs on an individual

[27] The parameter estimates are available on request.

level that was endogenous to the demand side error term, which would make endogeneity less of a concern. Fortunately, in the 2007 wave of the data, respondents were asked to report the price of the most recent purchase (and the quantity purchased) and the quality of marijuana purchased. As these are individual prices reported by quality type they are less likely to be correlated with the error term. Unfortunately reported prices are only available in one wave so we cannot use them for the entire analysis. However, the estimates using reported prices for the 2007 wave of the data are not significantly different from those reported in Table 4, hence, we are less concerned that price endogeneity is an issue once quality of marijuana is accounted for.

7 Conclusions

We present a model of marijuana use that disentangles the impact of limited accessibility from consumption decisions based solely on preferences. We find that both play an important role and that individuals who have access to the illicit market are of specific demographics. Our results indicate that observables and unobservables from marijuana use and access are positively related and that the elasticities of legalization and price are all significantly different in the selection model relative to the standard approach. The selection model indicates demand is much more elastic with respect to price. Counterfactual results indicate that making marijuana legal and removing accessibility barriers would have a smaller relative impact on younger individuals but still a large impact in magnitude. The probability of use among underage youth would increase by 38% and the probability of use among individuals in their thirties and forties would more than double.

We found that prices would need to be more than ten-fold higher than current levels in order to keep the frequency of post-legalization underage use the same as pre-legalization use even if underage users would still face the same restrictions as they face for alcohol use. Increasing prices by ten-fold is not feasible given that we would expect most users to resort to the black market. Hence, our results indicate that, while teens respond to prices, an environment where use is legalized will see an increase in the probability of use of at least 5% and a most 67% among teens.

For a population the size of the US our results indicate that a policy of marijuana legalization would raise a minimum amount between $132 million to $2.5 billion annually depending on which taxation scheme used (assuming individuals consume at a lower level of frequency).

If individuals consume at the midpoint of the frequency intervals then tax revenues would increase to between \$557 million to \$10 billion annually depending on the taxing scheme. Our findings are consistent with those from a 2005 report funded by the Marijuana Policy Project (Miron, 2005) which estimates legalization would raise tax revenues of \$2.4 billion annually if it were taxed like most consumer goods and over \$6 billion annually if it were taxed similarly to alcohol or tobacco. If we assume that 10% of the sales will be lost to the black market, tax revenues would decline to between \$502 million to \$9.2 billion (at the midpoint of the frequency). If all users who currently use (i.e., those who are currently willing to do something illegal) would buy on the black market tax revenues would be between \$452 million and \$8 billion (at the midpoint of the frequency).

Our study provides insight on the potential impacts of legalizing marijuana use. This is a multi-dimensional issue and so there are other important points to consider in the legalization debate, which are not addressed in this study. One such important issue concerns the long-run implications of legalization. For example, social acceptance may change so we may see a change in use due to this long-run effect. Another example, is that we may also see a change in the quality of marijuana offered should it become legal. Finally, we do not examine the health implications of use or the potential effects on labor market outcomes. These are important topics for future research in this area.

References

Adams, B. and B. Martin. 1996. “Cannabis: Pharmacology and Toxicology in Animals and Humans.” *Addiction,* 91, 1585-1614.

Adda J., B. McConnell and I. Rasul. 2011. “Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment,” http://www.ucl.ac.uk/~uctpimr/research/depenalization.pdf.

Albert J. and S. Chib. 1993. “Bayesian Analysis via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts.” *Journal of Business and Economic Statistics*, 11, 1-15.

Australian Crime Commission Illicit Drug Data Report. 2000. Report Number 2000-01, Australian Crime Commission (ACC) URL: http://www.crimecommission.gov.au/publications/illicit-drug-data-report.

Australian Crime Commission Illicit Drug Data Report. 2003. Report Number 2003-04, Australian Crime Commission (ACC) URL: http://www.crimecommission.gov.au/publications/illicit-drug-data-report.

Australian Crime Commission Illicit Drug Data Report. 2006. Report Number 2006-07, Australian Crime Commission (ACC) URL: http://www.crimecommission.gov.au/publications/illicit-drug-data-report.

Becker, G., K. Murphy, and M. Grossman. 2006. “The Market for Illegal Goods: The Case of Drugs.” *Journal of Political Economy,* 114(1), 38-60.

Bretteville-Jensen, A. and L. Jacobi. 2011. “Climbing the Drug Staircase: A Bayesian Analysis of the Initiation of Hard Drug Use.” *Journal of Applied Econometrics*, 26 (7), 1157-1186.

Caulkins, J., B. Kilmer, R. MacCoun, R. Pacula R., and P. Reuter. 2011. “Design Considerations for Legalizing Cannabis: Lessons Inspired by Analysis of California’s Proposition 19,” *Addiction.*

Chaloupka, F., K. Cummings, C. Morley, and J. Horan. 2002. “Tax, Price and Cigarette Smoking: Evidence from the Tobacco Documents and Implications for Tobacco Company Marketing Strategies.” *Tobacco Control,* 11(Suppl I), i62–i72.

Chaloupka, F., and K. Warner. 2000. “The Economics of Smoking.” In The Handbook of Health Economics Volume 1, Part 2 (eds. Newhouse J.P. and Cuyler A.J.), Amsterdam: Elsevier, 1539-627.

Chib S. and E. Greenberg. 1998. “Analysis of Multivariate Probit Models.” *Biometrika*, 85, 347-361.

Chib S. and L. Jacobi. 2008. “Analysis of Treatment Response Data from Eligibility Designs. *Journal of Econometrics*, 144, 465-478.

Ching, A., Erdem, T. and M. Keane. 2009. "The Price Consideration Model of Brand Choice," *Journal of Applied Econometrics*, 24(3), 393-420.

Clements, K. 2006. "Pricing and Packaging: The Case of Marijuana." *Journal of Business,* 79(4), 2019-44.

Clements, K.and X. Zhao. 2009. Economics and Marijuana: Consumption, Pricing and Legalisation Volume 422, Cambridge and New York: Cambridge University Press.

Clerides, S. and P. Courty. 2010. "Sales, Quantity Surcharge, and Consumer Inattention," University of Cyprus, Department of Economnics Discussion Paper 2010-07.

Damrongplasit, K., C. Hsaio, and X. Zhao. 2010. "Decrimilization and Marijuana Smoking Prevalence: Evidence from Australia." *Journal of Business and Economic Statistics*, 28(3), 344-356.

DeSimone, J. 1998. "Is Marijuana a Gateway Drug?" *Eastern Economic Journal*, 24(2), 149-64.

Donohue, J., B. Ewing, D. Peloquin. 2011. "Rethinking America's Illegal Drug Policy," NBER Working Papers: 16776.

Eizenberg, A. 2011. "Upstream Innovation and Product Variety in the U.S. Home PC Market." SSRN Working Paper: 1760828.

Eliaz, K. and R. Spiegler. 2011. "Consideration Sets and Competitive Marketing." *Review of Economic Studies*, 78(1), 235-262.

Framework Convention on Tobacco Control. 2012. "Encouraging Consumption, Funding Organized Crime." http://global.tobaccofreekids.org/en/solutions/international_issues/illicit_trade_smuggling.

Galenianos, M., Pacula R., and N. Persico. 2009. "A Search-Theoretic Model of the Retail Market for Illicit Drugs," NBER Working Paper 14980.

Jiang R., P. Manchandab, and P. Rossi. 2009. "Bayesian Analysis of Random Coefficient Logit Models using Aggregate Data." *Journal of Econometrics,* 149, 136-148.

Kim, J., P. Albuquerque, and B. Bronnenberg. 2010. "Online Demand Under Limited Consumer Search," *Marketing Science,* 29(6), 1001-1023.

Miron, J. 2005. "The Budgetary Implications of Marijuana Prohibition." http://www.prohibitioncosts.org/mironreport/.

Miron, J. and J. Zwiebel. 1995. "The Economic Case Against Drug Prohibition." *Journal of Economic Perspectives,* 9, 175-92.

Office of National Drug Control Policy (ONDCP). 2004. The Economic Costs of Drug Abuse in the United States, 1992-2002. Publication 207303. Washington, D.C.: Executive Office of the President.

Pacula, R, M. Grossman, F. Chaloupka, J. O'Malley, L. Johnston, and M. Farrelly. 2000. "Marijuana and Youth," NBER Working Papers No. 7703.

Pacula, R., B. Kilmer, M. Grossman and J. Chaloupka. 2010. "Risks and Prices: The Role of User Sanctions in Marijuana Markets. " *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy* 10.

Poulsen, H. and G. Sutherland. 2000. "The Potency of Cannabis in New Zealand from 1976 to 1996." *Science and Justice*, 40, 171-176.

Pudney, S. 2010. "Drugs Policy: What Should we do About Cannabis?" *Economic Policy,* 25, 165-211.

Sickles R., and P. Taubman. 1991. "Who Uses Illegal Drugs?" *American Economic Review, Papers and Proceedings* 81(2), 248-251.

Sovinsky Goeree, M. 2008. "Limited Information and Advertising in the US Personal Computer Industry." *Econometrica,* 76(5), 1017-1074.

United Nations Office of Drugs and Crimes. 2012. "World Drug Report, 2012," http://www.unodc.org/unodc/en/data-and-analysis/WDR-2012.html.

Van Ours, J. 2003. "Is Cannabis a Stepping-Stone for Cocaine?" *Journal of Health Economics,* 22, 539-54.

Veugelers, R. and B. Cassiman. 2005. "R&D Cooperation Between Firms and Universities. Some Empirical Evidence from Belgian Manufacturing. *International Journal of Industrial Organization* 23(5–6), 355–379.

Williams, J. 2004. "The Effects of Price and Policy on Marijuana Use: What Can Be Learned from the Australian Experience?" *Health Economics*, 13, 123–137.

Williams, J., J. van Ours, and M. Grossman. 2011. "Why Do Some People Want to Legalize Cannabis Use?" NBER Working Paper No. 16795

For Online Publication:

A Estimation

A.1 Model Fitting for Probit Model with Selection

For the estimation of the Probit model for marijuana use with selection based on binary access via MCMC methods we introduce the latent continuous access and marijuana use variables $\{a^*_{im}\}$ and $\{u^*_{im}\}$ and use the common latent variable representation of the probit

$$a^*_{im} = \mu^a_{im} + \eta_{im} = \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma} + \eta_{im}, \;\; a_{im} = I[a^*_{im} > 0]$$

$$u^*_{im} = \widetilde{\mu}^u_{im} + \varepsilon_{im} = \tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \varepsilon_{im}, \;\; u_{im} = I[u^*_{im} > 0] \qquad if \quad a_{im} = 1$$

where for each sample subject $\tilde{\mathbf{h}}_{im}$ refers to the combined covariate vector for the access model containing intercept, individual attributes, market-specific variables influencing access, and $\tilde{\mathbf{x}}_{im}$ is the combined covariate vector for the net utility model that contains the price variables $\{p_i\}$, individual attributes, market specific variables, year fixed effects, and state fixed effects in addition to the intercept. We define the vector of model parameters as $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}, \rho)$. Under the assumption that $(\eta_{im}, \varepsilon_{im}) \sim N_2(0, \Xi)$, where Ξ is $2x2$ covariance matrix with 1 on the diagonal and ρ on the off-diagonal, the likelihood of the model augmented with the latent access and net-use variables, $f(\mathbf{a}, \mathbf{u}, \{a^*_{im}\}, \{u^*_{im}\} | \boldsymbol{\theta}, \mathbf{W}, \{p_i\})$ can be expressed as

$$\begin{aligned}
&\prod_{i:a_{im}=0} \mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, \mathbf{1}) \; I[a^*_{im} \;\; \leq \;\; 0]^{a_{im}} \\
&\prod_{i:a_{im}=1} \mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, \mathbf{1}) I[a^*_{im} \;\; > \;\; 0]^{1-a_{im}} \quad \times \{\mathcal{N}(u^*_{im}|\tilde{\mathbf{x}}'_{im}\,\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma} \;, 1-\rho^2) \\
&\qquad\qquad\qquad\qquad\qquad \times \; \left[I[u^*_{im} \leq 0]^{1-u_{im}}\right] + \; \left[I[u^*_{im} > 0]^{u_{im}}\right]\}
\end{aligned}$$

where the inclusion of the latent data has improved the tractability of the likelihood (Albert and Chib, 1993).

For the Bayesian analysis we proceed with the common assumption of normal independent priors for the slope coefficients and correlation coefficient. The latter is restricted to the region $R = -1 < \rho < 1$ to ensure the positive definiteness of $\boldsymbol{\Xi}$. The joint prior is given by

$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0)\; \mathcal{N}(\boldsymbol{\gamma}|\mathbf{g}_0, \mathbf{G}_0)\; \mathcal{N}(\rho|r_0, R_0) \times R \tag{14}$$

The prior means are set at zero. In combination with large prior variances this implies relatively uninformative prior assumptions. It should be noted that in the context of our very large data set the influence of the prior is very small as the information from the data via the likelihood will dominate the inference about the model parameters summarized in the posterior distribution. The posterior distribution, with the parameter space augmented by the latent access and marijuana variables, $\pi(\boldsymbol{\theta}, \mathbf{a}^*, \mathbf{u}^*|\mathbf{a}, \mathbf{u})$, is proportional to product of the likelihood and the prior. We employ a straight forward Metropolis within Gibbs simulation algorithm with five blocks to generate draws from the posterior distribution of the parameter vector, as well as the marginal distributions of each parameter. By augmenting the parameter space with the latent access and net-use variables, the priors on slope coefficients are conditionally conjugate, thus allowing for normal updates of slope the coefficients. The latent variables are also normal updates. A Metropolis Hastings update is used for the correlation parameter as the structure of the covariance matrix and the likelihood do not allow a Gibbs update. The detailed steps of the algorithm are as follows:

First, we draw a^*_{im} from $\mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, 1)\ I[a^*_{im} < 0]$ for $i \in I_0$ and from $\mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma} + \rho(u^*_{im} - \widetilde{\mu}^u_{im}), 1-\rho^2)\ I[a^*_{im} \geq 0]$ for those subjects with $i \in I_1$.

In the second step, we draw u^*_{im} for all subjects $i \in I_1$ from $\mathcal{N}(u^*_{im}|\tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}), 1-\rho^2)\ I[u^*_{im} \leq 0]$ if $u_{im} = 0$ or from $\mathcal{N}(u^*_{im}|\tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}), 1-\rho^2)\ I[u^*_{im} > 0]$ if $u_{im} = 1$.

In the third step, we draw $\boldsymbol{\gamma}$ from $\mathcal{N}(\hat{\boldsymbol{\gamma}}, \hat{\mathbf{G}})$ with

$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}[\mathbf{G}_0^{-1}\mathbf{g}_0 + \sum_{i\in I_0} \tilde{\mathbf{h}}_{im} a^*_{im} + \sum_{i \in I_1} \tilde{\mathbf{h}}_{im}(1-\rho^2)^{-1}(a^*_{im} - \rho(u^*_{im} - \tilde{\mathbf{x}}'_{im}\boldsymbol{\beta})]$$

$$\hat{\mathbf{G}} = [\mathbf{G}_0^{-1} + \sum_{i\in I_0} \tilde{\mathbf{h}}_{im}\tilde{\mathbf{h}}'_{im} + \sum_{i\in I_1} \tilde{\mathbf{h}}_{im}(1-\rho^2)^{-1}\tilde{\mathbf{h}}'_{im}]^{-1}$$

where $i \in I_0$ refers to the subset of subjects with no access and $i \in I_1$ to those with access. In the fourth step we draw $\boldsymbol{\beta}$ based on the subjects in I_1 from $\mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{B}})$ where

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{B}}[\mathbf{B}_0^{-1}\mathbf{b}_0 + \sum_{i\in I_1} \tilde{\mathbf{x}}_{im}(1-\rho^2)^{-1}(u^*_{im} - \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}))]$$

$$\hat{\mathbf{B}} = [\mathbf{B}_0^{-1} + \sum_{i\in I_1} \tilde{\mathbf{x}}_{im}(1-\rho^2)^{-1}\tilde{\mathbf{x}}'_{im}]^{-1}.$$

In the last step we update ρ in Metrolpolis Hastings step based on the subjects in I_1, since the conditional posterior distribution of ρ is not tractable. Following Chib and Greenberg (1998) we generate proposal values for ρ' from a tailored student-t density $t_\nu(\mu, V)$ where μ

is the mode of

$$ln(\prod_{I\in I_1}\mathcal{N}(a^*_{im},u^*_{im}|W_{im}\boldsymbol{\delta},\Xi)\ , \quad \text{where} \quad \mathbf{W}_{im}=\begin{pmatrix}\tilde{\mathbf{h}}'_{im}\\ \\ \tilde{\mathbf{x}}'_{im}\end{pmatrix},\ \boldsymbol{\delta}=\begin{pmatrix}\boldsymbol{\gamma}\\ \\ \boldsymbol{\beta}\end{pmatrix} \ \text{and}\ \Xi=\begin{pmatrix}1 & \rho\\ \\ \rho & 1\end{pmatrix}$$

and V is the inverse of the Hessian of the density evaluated at μ. The proposed value ρ' is accepted with probability

$$\alpha=\min\left(1\ ,\ \frac{\pi(\rho')\ \prod_{I\in I_1}\mathcal{N}(a^*_{im},u^*_{im}|\mathbf{W}_{im}\boldsymbol{\delta},\Xi')\ t_\nu(\rho|\mu,V)}{\pi(\rho)\ \prod_{I\in I_1}\mathcal{N}(a^*_{im},u^*_{im}|\mathbf{W}_{im}\boldsymbol{\delta},\Xi)\ t_\nu(\rho'|\mu,V)}\right).$$

We repeat the above steps for M iterations after allowing for a burn-phase of M_0 iterations to allow for the convergence of the chain. We obtain a vector of M draws for each model parameters that reflects the posterior distribution of each parameter. In the main text we provide summaries of the posterior distributions in terms of the posterior means (coefficient estimate) and standard deviations or the 95% credibility intervals.

Finally, we address the issue of the unobserved prices as described in Section 2 by drawing the individual price for each subject i in market m with access from the market specific price distribution (10) at the beginning of each iteration of the MCMC algorithm.

A.2 Marijuana Use Prediction Counterfactual

We report the probabilities of marijuana use for various counterfactual scenarios in the paper based on our estimated probit model with selection. The probabilities are obtained using the standard Bayesian approach for prediction, allowing us to both use all the information from the parameter estimation summarized in the posterior distribution and to compute credible intervals. Let $n=1$ refer to a random subject from the sample with demographic characteristics and market features $\tilde{\mathbf{x}}_{n+1,m}$ for whom we want to predict the probability of marijuana use given the observed data $\Pr(u_{n+1,m}=1|\mathbf{a},\mathbf{u})$. Using the model on marijuana use with selection and the information on the model parameters from the posterior distribution we can obtain the probability of marijuana use based on following integral expression

$$\Pr(u_{n+1,m}=1|\mathbf{a},\mathbf{u})=\int\Phi(m_{n+1,m})\,\pi(\boldsymbol{\theta}\ |\mathbf{a},\mathbf{u})\ p(\tilde{\mathbf{x}}_{n+1,m},\tilde{\mathbf{h}}'_{n+1,m}|\mathbf{a},\mathbf{u})\,d\boldsymbol{\theta}\ \ d\,\tilde{\mathbf{x}}_{n+1,m},d\,\tilde{\mathbf{h}}_{n+1,m}$$

where $m_{n+1,m}=\ \tilde{\mathbf{x}}'_{n+1,m}\ \boldsymbol{\beta}\ +\rho\eta_{n+1,m}$ is the mean of marijuana use conditional on observed access. The term $\rho\eta_{n+1,m}$ accounts for selection on unobservables, where $\eta_{n+1,m}$ is found using information on the distribution of unobservables in the data exploiting information on the observed access. If $a_{n+1,m} = 0$, then it follows directly from the model of access that $\mu^a_{n+1,m} + \eta_{n+1,m} < 0$, or $\eta_{n+1,m} < -\mu^a_{n+1,m}$ where $\mu^a_{n+1,m} = \tilde{\mathbf{h}}'_{n+1,m}\boldsymbol{\gamma}$. As $\eta_{n+1,m}$ follows a standard normal distribution we generate $\eta_{n+1,m}$ from $TN_{(-\infty,-\mu^a_{n+1,m})}(0,1)$ for the subject. Similarly, for the case of $a_{n+1,m} = 1$ we have $\mu^a_{n+1,m} + \eta_{n+1,m} > 0$, so that we generate $\eta_{n+1,m}$ from $TN_{(-\mu^a_{n+1,m},\infty)}(0,1)$. Note that if we set $\rho\eta_{n+1,m} = 0$ we would implement the prediction based on the marginal model for marijuana use. While predictions are often based on the marginal model, this approach would lead us to considerably underpredict the benchmark case of use under pre-legalization relative to the observed use pre legalization, due to ignoring the important role that selection of unobservables play in the context of marijuana use and more generally in the use of illicit drugs. The $\pi(\boldsymbol{\theta} \mid \mathbf{a}, \mathbf{u})$ refers to the posterior distribution of the parameters and $p(\tilde{\mathbf{x}}_{n+1,m}, \tilde{\mathbf{h}}'_{n+1,m}|\mathbf{a}, \mathbf{u})$ to the empirical distribution of the individual, demographic and market characteristics in the sample data. The above integral expression can be estimated using the draws from the posterior distribution from the MCMC algorithm discussed in the previous section. Essentially, at each iteration of the MCMC algorithm after the burn-in phase, vectors $\tilde{\mathbf{x}}_{n+1,m}$ and $\tilde{\mathbf{h}}'_{n+1,m}$ are drawn from the data and $\Phi(m_{n+1,m})$ is computed using the current MCMC draws on the model parameters. All these draws give as a predictive distribution of the probability of use and we report the mean probability and the 95% credibility interval.

For the predicted probabilities of use by various demographic groups, we randomly draw $\tilde{\mathbf{x}}_{n+1,m}$ and $\tilde{\mathbf{h}}'_{n+1,m}$ from the corresponding subsample. Note that by predicting use for any random subject in the sample we assume that marijuana is freely accessible. To implement the counterfactual under legality of use and different price scenarios we set "Use if Legal" to one and adjust the price in $\tilde{\mathbf{x}}_{n+1,m}$. For counterfactuals using the pre-legal price we draw the price from the empirical distribution at each iteration. We also implement the prediction for the benchmark case of pre-legalization following the described approach by setting $\Phi(m_{n+1,m}) = 0$ if $a_{n+1,m} = 0$.

A.3 Model Fitting for Ordered Probit Model with Selection

We also estimate an ordered probit model with selection for the discrete ordered marijuana use variable, $y_{im} = 0, 1, 2$, for the analysis of tax revenues in section 5.3:

$y^*_{im} = \tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \nu_{im}$, where $y_{im} = 0$ if $y^*_{im} \leq 0$, $y_{im} = 1$ if $0 < y^*_{im} \leq \tau$ and $y_{im} = 2$ if $\tau < y^*_{im}$

where τ refers to the cut-off point that has to be estimated. The first cut-off point has been set to zero for identification purposes. The model for access remains unchanged and as before we assume that access and marijuana use may both be affected by unobserved factors so that $(\eta_{im}, \nu_{im}) \sim N_2(0, \Xi)$ where Ξ is $2x2$ covariance matrix with 1 on the diagonal and ρ on the off-diagonal. The likelihood of the model, $f(\mathbf{a}, \mathbf{y}, \{a^*_{im}\}, \{y^*_{im}\}|\boldsymbol{\xi}, \mathbf{W}, \{p_i\})$ where $\boldsymbol{\xi} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \rho, \tau)$, can be again expressed in terms of the latent data to improve the tractability of the likelihood (Albert and Chib, 1993) as

$$\begin{aligned}
\prod_{i:a_{im}=0} \mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, \mathbf{1})\ I[a^*_{im} &\leq 0]^{a_{im}} \\
\prod_{i:a_{im}=1} \mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, \mathbf{1}) I[a^*_{im} &> 0]^{1-a_{im}}\ \{\mathcal{N}(y^*_{im}|\tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}\ , 1-\rho^2)\ \times \\
(I[y_{im} &= 0]\ I[y^*_{im} \leq 0] + I[y_{im} = 1]\ I[0 < y^*_{im} \leq \tau] +\ I[y_{im} = 5]\ I[y^*_{im} > \tau])\}
\end{aligned}$$

We again assume independent normal priors for $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \rho)$ as in the probit model with selection. For the cut-off points it is sufficient to assume a priori that $0 < \boldsymbol{\tau}$.

To simulate the posterior distribution $\pi(\boldsymbol{\xi}, \mathbf{a}^*, \mathbf{y}^*|\mathbf{a}, \mathbf{y})$ we employ a 6 step MCMC algorithm that is an extended and modified version 5 step algorithm for the Bivariate Probit with Selection discussed above. We add a 6th step to draw the cut-off point and also adjust the generation of the latent utility $\mathbf{y}^*$. For the latter, we draw y^*_{im} for all subjects $i \in I_1$ from the truncated normal distributions $\mathcal{TN}_{(a,b)}(y^*_{im}|\tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\gamma), 1-\rho^2)\ I[a < y^*_{im} \leq b]$, where $(a = -\infty, b = 0)$ for $k = 0$, $(a = 0, b = \tau)$ for $k = 1$ and $(a = \tau, b = +\infty)$ for $k = 2$. To update the cut-off point we employ a Metropolis Hastings algorithm as the conditional posterior distribution is of an unknown form. To improve the performance we update the cut-off point marginalized over the latent utilities $\{y^*_{im}\}$ and generate the proposal values from the tailored student-t density $q(\tau) = t_{10}(\mu, V)$, where here μ is the mode of the likelihood of the access subjects with with y_{im}=1 and y_{im} =2, $f(\mathbf{a} = \mathbf{1}, \{a^*_{im}\}, \{y_{im}=1\}, \{y_{im}=2\}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \rho, \mathbf{W})$ and V is the inverse of the Hessian of the density evaluated at μ. We maximize the proportional conditional likelihood expression (omitted $\mathcal{N}(a^*_{im}|\tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma}, \mathbf{1})I[a^*_{im} > 0]^{1-a_{im}}$ as it does not depend on cut-off points)

$$ln\left(\prod_{I_1:\ y_{im}=1}\left[\Phi(\frac{\tau - m_{im}}{\sigma}) - \Phi(\frac{-m_{im}}{\sigma})\right]\right) + ln\left(\prod_{I_1:\ y_{im}=2}\left[1 - \Phi(\frac{\tau - m_{im}}{\sigma})\right]\right)$$

where $m_{im} = \tilde{\mathbf{x}}'_{im}\boldsymbol{\beta} + \rho(a^*_{im} - \tilde{\mathbf{h}}'_{im}\boldsymbol{\gamma})$ and $\sigma = \sqrt{1-\rho^2}$. The maximization is subject to the constraint that $\tau > 0$.

The proposed value τ , with $\tau > 0$, is accepted with probability

$$\alpha = \min\left(1\ ,\ \frac{f(\{y_{im}{=}1\},\{y_{im}{=}2\}|\{a^*_{im}\},\mathbf{a}=\mathbf{1},\boldsymbol{\gamma},\boldsymbol{\beta},\rho,\tau',\mathbf{W})\ t_\nu(\tau|\mu,V)}{f(\{y_{im}{=}1\},\{y_{im}{=}2\}|\{a^*_{im}\},\mathbf{a}=\mathbf{1},\boldsymbol{\gamma},\boldsymbol{\beta},\rho,\tau,\mathbf{W})\ t_\nu(\tau'|\mu,V)}\right)$$

where again we use the conditional form of the likelihood of marijuana use, omitting the marginal likelihood of access as it does not depend on the cut-off point. As in the algorithm for the probit model we draw the price for the access subject from the corresponding empirical distribution before each iteration of the algorithm. Table B1 presents the parameter estimates.

	Ordered Probit with Selection State Fixed Effects Specification			
	Cannabis Use		Access	
	Mean	Std Dev	Mean	Std Dev
Individual Attributes				
Male	0.301	(0.023) *	0.291	(0.012) *
Age in Teens Spline	0.090	(0.019) *	0.141	(0.017) *
Age in Twenties Spline	-0.021	(0.004) *	-0.039	(0.003) *
Age in Thirties Spline	-0.024	(0.004) *	-0.040	(0.002) *
Age in Forties Spline	-0.025	(0.004) *	-0.021	(0.002) *
Age over Forties Spline	-0.068	(0.007) *	-0.054	(0.003) *
Highest Education is High School			-0.009	(0.020)
Highest Education is Trade Degree			0.110	(0.016) *
Highest Education is University Degree			-0.021	(0.018)
Of Aboriginal Descent	0.188	(0.054) *	0.212	(0.047) *
In Good, Very Good, or Excellent Health	-0.292	(0.016) *		
Market and Policy Variables				
Price	-0.004	(0.002) *		
Use if Legal	0.274	(0.274) *		
Grams Possession is not Minor Offense	0.001	(0.000)		
Arrests Per Capita of Suppliers (Prevalence)			0.008	(0.068)
Live in City			-0.185	(0.013) *
High Potency	0.104	(0.161)		
Correlation (ρ)	0.038	(0.110)		
Number of Observations		51296		

Notes: Standard deviations in parentheses. * indicates 95% Bayesian confidence interval does not contain zero. Includes year fixed effects, state fixed effects, and a constant in access and use.

Table B1: Frequency of Use Estimates

B Price Counterfactual Calculation

For the counterfactual price calculation under the ordered probit model with selection we find the counterfactual price that implies a predicted "post-legal" probability of no use among teenagers under different scenarios (unrestricted access and legal, unrestricted access and illegal, restricted access and illegal for teenagers) that is comparable to the observed probability of no use before legalization in the data. Given the model we implement the analysis at the market level, finding the counterfactual prices for all teenagers in market m to match the observed probability of no use in their market, call this S_m^{obs}. Let $S_{im}^{Post}(\widehat{\Theta}, data; p_i)$ represent the probability of no use for teenager i in market m under a counterfactual of legalization, and p_i^{CF} the counterfactual price, where p_i^{CF} is chosen so that

$$\frac{\sum_{ni=1}^{n_m} S_{im}^{Post}(\widehat{\Theta}, data; p_i = p_i^{CF})}{n_m} = S_m^{obs}$$

where n_m refers to the number of teenagers in market m and information on the parameter set $\widehat{\Theta}$ comes from the estimated posterior distribution of the parameters. To find the counterfactual prices for each teenager in market m, we first predict post legalization probability of no use, as described in Appendix A.2, with price p_i coming from the corresponding (pre-legal) empirical distribution (10). If $S_{im}^{Post}(p_i) >= S_m^{obs}$, then we set $p_i^{CF} = p_i$. If a teenager's probability of no use is below the market average, we compute the counterfactual price, where p_i^{CF} is the price that equates

$$S_{im}^{Post}(\widehat{\Theta}, data; p_i^{CF}) = S_m^{obs}.$$

From our ordered probit model on marijuana use with selection it follows that the probability of no use is $S_{im}^{Post}(\widehat{\Theta}, data; p_i^{CF}) = \Phi(-(\mu_i = f(\widehat{\Theta}, data; p_i^{CF}))$, where μ_i is the conditional mean of marijuana use taking into account the selection of unobservables for teenager i so that under our model specification with price and age interactions

$$\Phi(-(\mu_i = p_i^{CF}\alpha_{1teen} + d_{i1}'\beta_1 + x_m'\beta_2 + L_m'\delta_1 + \rho\eta_{im})) = S_m^{obs}$$

where α_{1teen} is the coefficient on price for teenagers and $\rho\eta_{im}$ accounts for selection on unobservables with $\eta_{im} \sim TN_{(-\infty, -\mu_{im}^a,)}$ if $a_{im} = 0$ and $\eta_{im} \sim TN_{(-\mu_{im}^a, \infty)}$ if $a_{im} = 1$ as discussed in Appendix A.2. The counterfactual price is obtained from

$$p_i^{CF} = \frac{-\Phi^{-1}(S_m^{obs}) - d_{i1}'\beta_1 - x_m'\beta_2 - L_m'\delta_1 - \rho\eta_{im}}{\alpha_{1teen}}$$

where Φ^{-1} is the inverse of the normal CDF.

Exploiting the estimation results from the Bayesian approach, we employ the estimated posterior distribution for the model parameters to find the prices, rather than just the estimate in terms of the posterior mean. We simply use the all the M parameters draws from MCMC estimation algorithm described in Appendix A.3 and compute the counterfactual prices under each value of the draws obtaining the vector $(p_i^{CF^{(1)}}, ..., p_i^{CF^{(M)}})$. If the computed counterfactual price is negative which may occur as some of the support of the posterior distribution of the price coefficient for teenagers is over the positive real line, we set the price to zero. In the paper we report the counterfactual price averaged over all the individuals and over all draws. We implement the counterfactual analysis under different settings. For example, for the setting when access is restricted and use considered illegal for teenagers, we set the probability of no use for a teenagers that reported no access equal to one and proceed with the above analysis as described above for teenagers that reported access. The average market probability to match (by the access subjects), S_m^{obs}, is now based on access subject pre-legalization. To assume legalization for teenagers, we simple set the "Use if Legal" variable to one for all teenagers in the above described steps.