More than Joints: Multi-Substance Use, Choice Limitations, and Policy Implications

Michelle Sovinsky, Liana Jacobi, Alessandra Allocca, and Tao Sun*

December 1, 2023

As illicit substances move into the legal product space, substitution patterns with legal products become more salient. In particular, marijuana legalization may have implications for the use of other legal "sin" goods. We estimate a structural model of multi-product use of illegal and legal substances considering joint use, limited access to illicit products, and persistence in use. We focus on a young person's choice to consume marijuana, alcohol or cigarettes (and possible combinations), and we find that sin goods are complements. Furthermore, our findings emphasize the necessity of accounting for joint consumption and access to obtain correct price sensitivity estimates. Post-legalization, youth marijuana use would increase from 25% to 37%. However, counterfactual results show that a combination of (reasonable) tax increases on all goods along with enforcement against illegal use can potentially revert use to pre-legalization levels. The earlier the tax increases are implemented the more effective they are at curbing future use. Our results inform the policy debate regarding the impact of marijuana legalization on the long-term use of sin goods.
JEL Classification: C11, D12, L15, K42, H2, L66, C35
Keywords: complementarity, marijuana legalization, limited choice sets, data restrictions, discrete choice models

---

[0] *Sovinsky (corresponding author) is at the University of Mannheim, Department of Economics and CEPR, michelle.sovinsky@gmail.com; Jacobi is at the University of Melbourne, Department of Economics, ljacobi@unimelb.edu.au; Allocca is at the University of Munich, Department of Economics, alessandra.allocca@econ.lmu.de; Sun is at the University of Melbourne, Department of Economics, tao.sun1@unimelb.edu.au;. Sovinsky acknowledges support from the European Research Council Grant #725081 FORENSICS and from the German Research Foundation (DFG) through CRC TR 224 (Project A02). Allocca acknowledges support from the German Research Foundation (DFG) through CRC TR 190 (Project B04). Sun acknowledges the financial support by the Australian Government Research Training Program (RTP) Scholarship and the Henry Buck Scholarship. We are grateful to Boyoung Seo, Laura Grigolon, Ali Yurukoglu, Mo Xiao, Steve Tadelis, Bob Miller and seminar participants at APIOC (Sydney), Bristol, Exeter, Frankfurt Goethe, IIOC, LMU, Macci Summer Institute (Mannheim), Frontiers in Empirical IO Workshop (Mannheim), Melbourne, Mines Paris Tech, Paris PSE, Penn State-Cornell Workshop, and Toulouse for helpful comments.

1 Introduction

The option to buy marijuana legally is rapidly becoming the norm. As of November 2023, selling and buying legally is possible in 24 U.S. states as well as in Uruguay, Canada, and Thailand. It is also on the agenda in Germany, Italy, Czech Republic and the UK with more US states poised to put it to a vote.[1] The arguments in favor of legalization are multi-faceted and include increased tax revenues, health benefits, reducing the black market including eliminating pathways to harder drugs, and the ability to regulate consumption.[2]

While these elements have differing potential to be realized, there is one point that is often overlooked - marijuana is rarely used in isolation. Most consumers who report using marijuana use it together with other “sin” goods (e.g., alcohol and cigarettes). This is also true of the youth. According to the Monitoring the Future survey of high school students, approximately 13% reported they used marijuana and alcohol in the same occasion over the past year. This suggests that regulation targeting alcohol and cigarettes could potentially influence use of marijuana. At the very least it implies we should look more closely at multiple use among sin goods when assessing the impact of marijuana legalization. Furthermore, long-run changes in behavior may become more permanent due to the presence of addictive elements (as with cigarettes and alcohol). These are critical concerns to weigh, especially regarding the impact on the youth.

In this paper we assess the impact of marijuana legalization on consumption of sin goods. We focus on a young person's choice to consume marijuana, alcohol or cigarettes (and possible combinations) within a framework of multi-substance use. Specifically, we estimate a structural model of multi-product use of illegal and legal substances that incorporates bundle use and complementarities among products, accounts for limited access to illegal marijuana, and allows for persistence in use.

---

[1] Use is legal in Alaska, Arizona, California, Colorado, Connecticut, Delaware, Illinois, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nevada, New Jersey, New Mexico, New York, Ohio, Oregon, Rhode Island, Vermont, Virginia, and Washington. Uruguay became the world's first country to legalize marijuana in 2013, Canada followed in 2018, and the latest to legalize was Thailand in 2022. Use is also legal in the countries of Malta, Mexico, South Africa and Georgia, and Luxembourg but selling remains illegal (outside of cannabis clubs).

[2] For a discussion on the policy landscape in the U.S., see Haffajee and Mauri (2021).

We estimate our model using data on youth use in the US from two datasets - Monitoring the Future (MTF) and the Panel Study of Income Dynamics (PSID) during a period when recreational marijuana was not legal anywhere in the U.S. (2004-2013). Neither survey is ideal: the MTF does not track youth over time and the PSID does not contain information on access to illegal drugs. Regarding the latter, Jacobi and Sovinsky (2016) show it is important to control for limited access to marijuana (due to the illegal nature of the product) in order to obtain correct estimates of substitution patterns. To overcome limitations in both data surveys, we propose a method to combine the data to take advantage of access information (and larger sample sizes) from the MTF and information on persistence in use from the PSID. The conceptual idea is to simulate an MTF individual's past use behavior based on an empirical distribution of lagged use generated from the PSID. The empirical past-use distribution is linked to an MTF cohort based on demographics. Our methodology is in the spirit of Berry, Levinsohn, and Pakes (1995) which draws a "consumer" from an empirical distribution generated from another dataset. Our approach allows us to turn cross-sectional data into a "matched panel" to overcome data limitations by using information from more than one source to inform our estimates. The approach (which we refer to as a matched panel) is novel in the literature and is distinct from a pseudo panel, where the latter is formed from repeated cross-sections on the basis of combinations of demographic characteristics.[3]

To implement the empirical analysis, we proceed within the Bayesian multinomial probit (MNP) framework. We introduce the first MNP framework for bundle choice models and show how the framework can incorporate unobserved bundle taste correlation (as in Gentzkow (2007)) and, additionally, take into account limited choice sets and persistence via the matched panel approach. We propose an efficient and robust MCMC-based inferential strategy for the model estimation and the counterfactual policy analysis.

The results suggest that marijuana, cigarettes, and alcohol are complements in consumption. The results also show that it is important to control for access, which reveals that consumers are less sensitive to price changes than a full access framework indicates. In addition,

[3] Although, our matched panel approach is related in the sense that we draw a consumer of a similar demographic group that we match across the two datasets. The seminal papers are Deaton (1985) and Moffitt (1993).

we find that we would incorrectly measure complementary effects among products (and in some cases conclude they were substitutes) if we did not allow for bundle consumption of multiple substances. Finally, there is persistence in use for all substances.

We find that marijuana legalization will impact use of all substances, where use of marijuana among youth would increase from 25% to 37% under legalization. We show that 17% of the increase in use is due to decreased police enforcement, and that 28% is due to decreased access barriers. However, the impact of legalization can be mitigated through taxing not only marijuana, but also alcohol and cigarettes. In fact, there are (reasonable) combinations of tax increases that will compensate for the increase in marijuana use due to legalization (when used in combination with continued police enforcement). For example, Nevada was the only state to increase tax on at least one sin good post-legalization (alcohol). Our results imply, had they also increased cigarette taxes by a mere 8% (coupled with a 8% higher marijuana tax), youth use would have been close to pre-legalization levels. It is notable that none of the jurisdictions implemented an increase in taxes on both sin goods when legalizing marijuana. Our findings provide previously unexploited tools to control youth use post-legalization.

Finally, our results show that the timing of tax increases is critical due to persistence in use. The earlier post-legalization the tax increases are implemented the more effective the policies are at curbing use in the future.

Our paper is related to several strands of literature. There is a large literature on substance use, although the majority consider use among one or two substances in a static setting. Those that study the use of all three substances (although not simultaneous use) include Cameron and Williams (2001), Zhao and Harris (2004), Clements, Lan, and Zhao (2010) and Miller and Seo (2021). Cameron and Williams (2001) estimate the price responsiveness of participation in cannabis, alcohol and tobacco use in a static framework, where the products are related to each other through shared observables. Zhao and Harris (2004) and Clements et al. (2010) estimate a static multi-variate model which incorporates unobserved correlation across choices as well as observed correlation via shared regressors in each choice equation. Miller and Seo (2021) consider the substitution between legal cannabis products, alcohol and tobacco using a flexible demand system for legal substances.[4] Finally, Deza (2015) estimates a dynamic model of pair-wise choices among alcohol, marijuana and hard drug use that allows for joint use of the paired substances. Her paper is focused on the gateway effect of drug use, so she includes covariates based on treatment rather than market or product-specific variables, such as substance prices or quality.[5] These papers have different estimation techniques, consider different time periods and geographic regions, even so, most find evidence that marijuana and tobacco are complements. However, the findings on the relationship between marijuana and alcohol are more mixed. Our work differs from these papers in several important ways. First, we allow for bundle choices. Second, as one of the products is an illicit substance, we control for limited choice sets arising from restricted access to marijuana. Finally, we incorporate persistence in a bundle framework.

Our results also inform the growing literature on the market effects of recreational marijuana legalization. Recent papers focus on the effects of taxation (Hansen, Miller, and Weber (2017); Hollenbeck and Giroldo (2022); Hollenbeck and Uetake (2021); Jacobi and Sovinsky (2016)), market entry restrictions (Thomas (2019)), cross-border sales (Hansen, Miller, and Weber (2020)), and cross-border externalities in use (Hinnosaar, Liu, and Loaeza-Albino (2023)).[6] We will add to these by examining the impact of taxing sin goods on marijuana use as well as non-tax policy tools.

Our paper is also related to the literature that measures complementarities between other products allowing for bundle use. The seminal paper in this area is Gentzkow (2007) who studies complementarity among newspapers. Others include Liu, Chintagunta, and Zhu (2010) (broadband and related categories), Thomassen, Smith, Seiler, and Schiraldi (2017) (grocery items), Ershov, Laliberté, and Orr (2018) (soft drinks and potato chips), Fosgerau, Monardo, and De Palma (2020) (RTE cereal brands) and Yu (2022) (yogurt brands).[7]

---

[4] Ritter and Sotade (2017) show a negative association between alcohol, tobacco and marijuana consumption in Australia since the 90s, pointing to potential complementarities among the substances.

[5] Deza (2015) also estimates a trivariate logit model but does not estimate the full model that allows for all eight possible bundles of marijuana, alcohol and hard drugs.

[6] Theoretically, Arnabal Rocca (2022) and Auriol, Mesnard, and Perrault (2023) study the optimal government intervention and the policy-mix to reduce consumption when marijuana is legalized.

[7] Iaria and Wang (2021) provide instrument-free identification and estimation methods that solve the

Our paper is also related to the literature on limited choice sets. These include Sovinsky Goeree (2008), Ching, Erdem, and Keane (2009), Ching and Hayashi (2010), Kim, Albuquerque, and Bronnenberg (2010), Eliaz and Spiegler (2011), Clerides and Courty (2017), and Jacobi and Sovinsky (2016). The latter paper examines the impact of legalization on use while controlling for limited access to marijuana. This current work differs from Jacobi and Sovinsky (2016) in three important ways. First, we examine choices of multiple substances. Second, we control for persistence in use decisions. Finally, we consider underaged individuals in the U.S.

Starting from the theoretical framework of Becker and Murphy (1988), several papers have empirically analyzed the impact of past behavior on the current use of sin goods (e.g., Becker, Grossman, and Murphy (1994); Chaloupka (1991); Gilleskie and Strumpf (2005); Grossman, Chaloupka, and Sirtalan (1998), and more recently Darden (2017); Deza (2015); Hai and Heckman (2022)). We contribute to this literature by focusing on the potential complementarities among licit and illicit substances accounting for limited choice sets.

Finally, our work is related to Bayesian methods for estimating structural demand models. This literature includes Jiang, Manchanda, and Rossi (2009), who use random coefficient logit models; Imai, Jain, and Ching (2009) who propose a dynamic discrete choice model, and Dubé, Hitsch, and Rossi (2010) who study consumer inertia.

The paper is structured as follows. We provide background on the legalization status of marijuana and discuss the data in section 2. We introduce our model in section 3. In section 4, we provide the econometric methodology including our new matched panel data augmentation approach and discuss identification. We present the results and counterfactual policy experiments in sections 5 and 6, respectively. We discuss goodness-of-fit and robustness of the estimates in section 7. Finally, we conclude.

challenge of dimensionality related to bundling.

2 Background and Data

The first licensed sales of recreational marijuana in the US took place in 2014 in Colorado and Washington state.[8] As of this writing, 24 states permit legal sales, but sales remain illegal nationwide. We wish to examine the impact that a nationwide legalization policy would have on consumption of a vulnerable group, the youth. To cleanly implement this, we use data on choices of individuals before it was possible to buy legally in any state. Therefore, the last year in our sample is 2013 which is a year before the first legal sales.

Our data start in 2004 and consist of individual cross-sectional survey data from Monitoring the Future (MTF), which is a survey of US high-school students.[9] These data are particularly useful as they contain consumption information for marijuana, alcohol and cigarettes, as well as variables related to marijuana accessibility and police enforcement. We complement these with data on substance prices and marijuana quality obtained from administrative, crowd-sourced, and industry sources. Finally, we use panel data on (past) substance consumption from the Panel Study of Income Dynamics (PSID). We discuss these in turn.

2.1 Monitoring the Future Data

MTF data are collected annually, and we use 10 waves from 2004 to 2013 across four US regions (North East, North Central, South, West). Respondents are students in the last year of high school (aged between 17 and 19).[10] After dropping individuals with missing observations our sample consists of 109,747 individuals. Table 1 shows summary statistics.

---

[8] Colorado amendment 64 was passed on November 6, 2012, which led to legalization in January 2014. Washington Initiative 502 was passed on December 6, 2012, which led to legalized sales in July 8, 2014.

[9] We should note that while we could not get access to panel level MTF data, they do exist, and are available to researchers who meet certain criteria (including geographical location).

[10] The survey has been conducted using a multistage random sampling procedure. In Stage 1, there is the selection of particular geographic areas, Stage 2 is the selection (with probability proportionate to size) of one or more high schools in each area, and Stage 3 is the selection of 12th graders within each high school. The samples for the MTF study are representative of high school seniors throughout the 48 US contiguous states (MTF Codebook, paragraph "Representativeness and Validity"). Alaska, Hawaii, and the District of Columbia are not included in the sample. Moreover, Dills, Goffard, and Miron (2017) report that Oregon does not appear in the years 2011-2013.

Nearly half of the individuals are male and more than 60% are non-Hispanic whites. A large proportion of them are still in school, so instead of controlling for years of education (as it is ongoing) we control for parental education. About 52% have at least one parent with a university degree or higher, while 27% have parents with a maximum high-school degree. Approximately 80% of the students live in an area with at least 50,000 residents (SMSA).

Table 1: MTF Summary Statistics

	Year									
	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Demographics										
Aged 18 or Older	0.54	0.54	0.55	0.56	0.57	0.56	0.56	0.56	0.56	0.57
Male	0.48	0.48	0.48	0.48	0.49	0.48	0.49	0.50	0.50	0.50
Non-Hispanic White	0.68	0.69	0.66	0.64	0.62	0.61	0.59	0.58	0.62	0.60
Parents Max High School Education	0.28	0.27	0.27	0.28	0.26	0.26	0.27	0.26	0.25	0.26
Parents Max University Education	0.52	0.53	0.52	0.51	0.51	0.49	0.50	0.52	0.54	0.51
Live in a SMSA	0.82	0.79	0.78	0.80	0.82	0.82	0.82	0.80	0.80	0.82
Marijuana Use										
Current User (Used in Last 30 Days)	0.23	0.23	0.22	0.23	0.23	0.25	0.26	0.27	0.27	0.26
Cigarette Use										
Current User (Used in Last 30 Days)	0.25	0.24	0.23	0.23	0.22	0.21	0.20	0.19	0.18	0.16
Alcohol Use										
Current User (Used in Last 30 Days)	0.51	0.50	0.49	0.48	0.47	0.47	0.46	0.44	0.46	0.43
Number of Observations	*11,405*	*11,432*	*11,122*	*11,424*	*10,875*	*10,641*	*11,194*	*10,941*	*10,903*	*9,810*

The survey asks several questions regarding the consumption of marijuana, alcohol and cigarettes.[11] If an individual reports consuming a substance on one or more occasions in the previous month, we define the individual as a current user. As Table 1 shows, the average proportion of current marijuana users is between 23% and 27%. Alcohol use is declining over the period from 51% to 43%.[12] Likewise, cigarette use is declining (from 25% to 16%) and is used at the lowest frequency among all substances.

Multi-substance Consumption Table 2 presents statistics on joint substance use, where we define multi-substance use as use of multiple substances in the past month. The majority who used alcohol reported using marijuana (10%) in the last month, or cigarettes (6%) or both (11%). Less than half of the individuals who drank alcohol used this substance alone (20% in Table 2 compared to 47% in Table 1). Multi-substance use is even more

[11] For the validity of MTF self-reporting on drug use, see http://www.monitoringthefuture.org/

[12] A similar trend has been reported by Kilmer et al. (2014), where they use data from the National Survey on Drug Use and Health (NSDUH).

pronounced for marijuana. For example, in 2004, according to Table 1, 23% reported using marijuana; however, Table 2 shows only 3% used it in isolation. Almost all marijuana users also used either alcohol or alcohol and cigarettes.

Table 2: MTF Multi-Substance Use

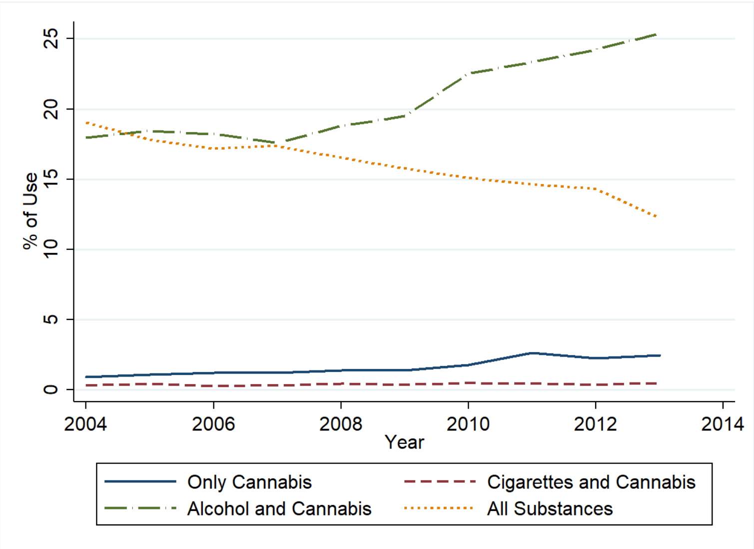
	Years		
Current User	**All**	**2004**	**2013**
Used No Substances	0.47	0.43	0.49
Marijuana Alone	0.03	0.02	0.04
Cigarettes Alone	0.03	0.03	0.02
Alcohol Alone	0.20	0.22	0.18
Marijuana and Cigarettes	0.01	0.01	0.02
Marijuana and Alcohol	0.10	0.08	0.12
Cigarettes and Alcohol	0.06	0.08	0.04
Used All Substances	0.11	0.13	0.08
Number of observations	*109,747*	*11,405*	*9,810*

Figure 1 depicts how the consumption of marijuana bundles has evolved over time. Notice that the use of marijuana alone or with alcohol increased between 2004 and 2013, while the joint use of all substances (dotted yellow line) declined. These simple descriptive statistics indicate the importance of considering consumption decisions for marijuana in combination with other substances.

One concern is that observing use of multiple substances in a month may not be an adequate measure of multi-substance use. The MTF data do not contain information on simultaneous use for all participants and all substances. However, a subsample of individuals (6,738 individuals) was asked whether they used marijuana and alcohol simultaneously ("so that their effects overlapped") in the past year. We cannot use these data to estimate the model, mainly because they only ask about one of the four potential bundles and only for the past year. However, the data are valuable because they allow us to gauge how well our definition of joint use captures simultaneous use.

Our definition of joint use and the simultaneous use question generate very similar use probabilities by demographics. For example, 58% of males reported they have used the two

Figure 1: Evolution of (Multi-) Substance Use



substances simultaneously in the past year and 57% of males reported joint use based on our definition. Likewise, among individuals aged 18 or older, 56% used them together and 54% reported joint use based on our definition. Similarly, our definition matches with true simultaneous use across races, but to a lesser extent (69% in the simultaneous use definition and 62% in our definition). In all, we take this as support that our definition of joint use reflects simultaneous use across the other potential bundles as well.

Marijuana Access Given that marijuana is illegal over our period, the decision to consume depends on whether an individual has access to the drug. As we discussed earlier, the MTF survey not only contains demographic, market, and illicit drug use information, but also information on accessibility to marijuana. We construct a variable that reflects access to marijuana using answers to two questions. The first question asks how easy it would be to obtain marijuana. If they answer it is "very/fairly easy" to get marijuana (about 85% of the sample) then we set our accessibility variable to one; if they report it is "fairly/very difficult" (about 9%) or "probably impossible" (about 5%) then we set access equal to zero. The second question asks how often they see drug dealers in their neighborhood. If they report they see them regularly (i.e. "one to three times monthly" (about 7%), "one to three times weekly" (about 7%), or "daily/almost daily" (about 8%)), we assume that they have access and set our access variable to one. However, if they never see drug dealers (61%) or see them "less than once a month" (about 16%) we assume they do not have access. Most individuals were not asked both questions, but each individual was asked at least one question. Our access definition implies about 77% of respondents had access to marijuana on average.[13] The unconditional probability of marijuana use is 25%, whereas the probability conditional on access is 32%. One important observation is that who has access differs by gender, race, age, and educational status of the parents. For example, those who have access are more likely to be white (64%) or over 17 (55%). This suggests that those who have access to marijuana are a selected sample, and indicates it is important to control for it. Our model addresses this issue.

2.2 Prices, Quality, and Police Enforcement

We use cigarette and alcohol taxes reported in administrative data. For cigarettes, we use the (weighted average) tax on a pack of cigarettes given in *The Tax Burden on Tobacco*. For alcohol, we use beer excise tax (normalized for a 6-pack of 12-ounce bottles) as reported by the Tax Foundation.[14]

We obtain marijuana prices and quality from industry publications and crowd-sourced data. Specifically, we use the price per gram reported by High Times and PriceOfWeed.com. High Times is a monthly magazine that publishes the prices of marijuana by strain. PriceOfWeed.com is a website that crowdsources the street value of marijuana from buyers.[15] Both sources are available for all states in the US, but neither covers the whole period of our data (2004-2013). High Times reports prices until 2011 and PriceofWeed.com starts in 2010. We use data from both sources where the overlap in years (2010 and 2011) shows consistent prices, which verifies our data do not have spurious price changes driven by moving from

[13] We provide robustness checks under alternative access definitions in section 7.

[14] The Tax Burden of Tobacco, Column 2 of Table 13B, 2014; https://taxfoundation.org/state-sales-gasoline-cigarette-and-alcohol-tax-rates). MTF reports on the region in which the individual lives so the region prices are state population-weighted prices.

[15] For High Times the relevant data are in the Section Trans High Market Quotations. PriceofWeed.com users (anonymously) report the amount they purchased, the price, and the quality of their most recent transaction.

one data source to another. We determine the price per gram by computing the weighted average over months and geographic location.

The major psychoactive chemical compound in marijuana is delta-9-tetrahydrocannabinol (or THC), which is responsible for the "high" caused from using marijuana. The high obtained varies depending on the quality of the strain, where higher potency demands a higher price on average. Therefore, it is important to control for the potency of the marijuana sold in the market. Transactions data from PriceOfWeed.com contains information on the quality of the purchase, where high-quality indicates the level of THC derived from the strain was of high potency. For the High Times data, we construct a quality measure based on the strain (which they report) by matching the strain to its level of THC.[16] According to Dutch Passion, one of the biggest suppliers of cannabis seeds worldwide, a THC level of 15% or above is classified as "very to extremely" high quality.[17] Indeed, strains with a THC level of 15% or above are typically associated with words like "potent" and "strong" or are said to have quick effects on the body. Hence, they are perceived by users to be of better quality as measured by the high-inducing effects. Therefore, we classify strains with a level of THC above 15% as high-quality and those with a level of THC below 15% as low-quality.[18] We aggregate the data from both sources to obtain an index of high-quality cannabis given by the percentage of high-quality marijuana purchases in each location, time period.[19]

Table 3 contains summary statistics for prices and marijuana quality for each region (defined by the Census Bureau) for the first and last year of our sample. Cigarette taxes on a package ranged from \$0.72 to \$1.98 in 2004 and from \$1.94 to \$4.15 nine years later, where

[16] We used information from multiple websites to match the strain to a THC level. These include (the portion before ".com") weedsmokersguide, marijuana-strains, organicann, dutch-passion, marijuana-seeds-weed, wikileaf, seedfinder, allbud, thebcsc, budderweeds, urbandictionary, leafly,cannasos, buddyboybrands, cannafo, cannabisreports, wikipedia, coloradocannabistours, libertycannabis, marijuanabreak, and natural-cannabis.

[17] https://dutch-passion.com/en/blog/what-is-a-high-amount-of-thc-for-cannabis-n969).

[18] For example, the strains skunk, haze, kind and crack are classified as high quality, while the strains schwag, reggie, commercial, ditch weed, dirt, brick bud, shake, wack, bunk and mids are classified as low quality.

[19] Out of 1,454 unique strains, we classify 1,153 strains while the remaining 301 are dropped from the sample as the amount of THC contained cannot be determined. These observations represent only the 0.3% of the entire sample. We provide more information in Appendix A.1.

Table 3: Region Descriptive Statistics

	North East		North Central		South		West	
	2004	**2013**	**2004**	**2013**	**2004**	**2013**	**2004**	**2013**
Variable								
Cigarettes Tax (Package)	1.98	4.15	1.23	2.63	0.72	1.94	1.25	2.08
Beer Tax (6-pack)	6.75	7.31	9.00	11.25	18.56	28.69	10.69	12.94
Marijuana Price (Gram)	11.43	11.58	10.84	11.21	10.72	10.94	10.80	8.00
Ratio High Quality Marijuana	0.60	0.49	0.63	0.50	0.74	0.52	0.82	0.47
Police Enforcement	0.49	0.49	0.56	0.57	0.51	0.53	0.46	0.46

Notes: Weighted (by population) average prices per dollar (cents) for cigarettes and marijuana (beer). Alaska, Hawaii, DC, and Oregon (for 2011-2013) are excluded. The mean for the variable Police Enforcement has been constructed from the answers to the MTF survey question: *Are you not using marijuana because you are concerned about being arrested? Yes/No.*

the North East region had the highest prices. The highest price for marijuana is also found in the North East. On average marijuana is around \$11 per gram, with a minimum price of \$8 in the West region in 2013.[20] Interestingly, the percentage of high-quality marijuana on the market is lower in 2013 relative to 2004. In contrast, beer taxes on a 6-pack are the lowest in the North East and the highest in the South.[21]

Given that marijuana is not legal (over our time period), one of the deterrents to using may be the fear of getting arrested. MTF elicits whether the concern of being arrested influenced a person's decision to use marijuana. The data indicate this is indeed important to control for as more than half of individuals answer in the affirmative, with the North Central Region exhibiting the highest percentage in 2013 (57%).[22] We use this variable as a proxy for perceived police enforcement in the area.

2.3 Panel Study of Income Dynamics

As noted previously, the available MTF data does not contain information on use across time. However, past use is important for understanding current choices, particularly when

[20] Using the same web-based source, Dills, Goffard, and Miron (2016) find that the average marijuana price in Colorado in 2013 was around \$7 per gram. This is consistent with the average price we find in the West region in 2013.

[21] Relative to the original source, beer taxes have been converted in cents and normalized for a 6-pack.

[22] This question is asked to a sample of individuals. For those who were not asked, we assign the average based on the time period, geographic location, and whether they live in an SMSA.

addictive elements play a role and we want to assess implications of policy changes beyond the short-term. Therefore, we augment the MTF with panel data on consumption decisions among similar individuals from the Panel Study of Income Dynamics (PSID)/Transition into Adulthood Supplement (TAS). The PSID/TAS is a panel dataset that collects information on adolescents between 17 and 27 years old. They are surveyed every odd year, and we use 5 waves from 2005 until 2013. Our initial sample consists of 6,398 observations for about 2,000 different individuals.[23]

Table 4: PSID Summary Statistics

			Year		
	2005	**2007**	**2009**	**2011**	**2013**
Demographics					
Age	19	20	21	22	23
Male	0.46	0.47	0.46	0.47	0.47
Non-Hispanic White	0.48	0.47	0.47	0.45	0.46
Max High School Education	0.85	0.81	0.76	0.72	0.66
Max University Education	0.02	0.08	0.15	0.20	0.27
Live in a Metropolitan Area	0.76	0.76	0.77	0.79	0.78
Marijuana Use					
Current User	0.31	0.26	0.27	0.29	0.30
Cigarette Use					
Current User	0.24	0.25	0.23	0.21	0.20
Alcohol Use					
Current User	0.62	0.61	0.64	0.64	0.71
Number of Observations	*695*	*1,078*	*1,493*	*1,769*	*1,363*

Table 4 presents descriptive statistics for the PSID data. As in the MTF data, the sample is almost evenly distributed between genders, although there is a slightly smaller percentage of non-hispanic whites (47%). As individuals age over time, the percentage of college graduates increases from 2% to 27%. Most of the individuals live in a metropolitan area (77%).[24] For the PSID data, if an individual reports consuming a substance on one or more occasions in the previous year, we define the individual as a current user. MTF and PSID respondents are similar with respect to marijuana use in terms of evolution, although the use probabilities are a bit lower (higher) for cigarette (alcohol) use. The data show a

[23] To be consistent with the MTF sample, we drop individuals from Alaska, Hawaii, DC and Oregon (the latter for the years 2011-2013) which are not contained in the MTF data. See the discussion in Section 2.1

[24] The Have Beale code (metropolitan/non-metropolitan area) is available in the PSID family survey and we use it to determine whether the individual lives in a more urban area.

stable trend for marijuana use and a downward trend for cigarette use, as in the MTF. We also see an upward trend for alcohol use as these individuals age with time.

The PSID shows persistence in use with 91% (for marijuana), 32% (for cigarettes) and 94% (for alcohol) of past users using the substance currently. We examined the data along several demographic and geographic dimensions. We found that gender, race, and age are the most important demographics associated with substance choices. The percentage of male individuals who consumed in the past and are current users is similar to that of females for marijuana (90% and 91%). This does not hold for cigarettes and alcohol: 33% of females smoked in the past and are current smokers relative to 30% of males; moreover, 93% of females drank in the past and currently drink relative to 96% of males.

The most notable difference is by race. Whites who have used any substance in the past period are more likely to be current users relative to non-whites. This finding is consistent across substances. For example, 93% of whites who consumed marijuana in the past are also current users, whereas for non-whites this percentage goes down to 89%. A similar difference in lag use holds for drinking (96% for whites and 92% for non-whites). Finally, 90% of individuals over 17 who have used marijuana in the past are also current users, 34% who have smoked in the past period are current smokers, and 94% who have used alcohol in the past period use the substance currently.

We construct an empirical distribution of lagged use that varies across groups based on gender, race, and age (over 17), which we use to construct the matched panel. Given the slight differences in use probabilities, we rescale PSID lagged use by group to reflect substance use rates in the MTF in the last month. We provide the empirical distribution by demographic group in the Online Appendix. After dropping the first panel observation, the PSID dataset consists of 2,751 observations.

3 Model of Multi-Substance Use and Access

We model an individual's choice to consume marijuana, cigarettes, or alcohol (and possible combinations) over time. We first discuss the model of multi-substance use and how we account for limited access. We then add persistence into the framework.

We present a model where an individual i chooses whether to consume good $j \in \{1, ..., J\}$ and whether to consume this product together with other products. Our model allows for complementarities among products by building on the work of Gentzkow (2007). Following Gentzkow (2007), we denote the set of consumption bundles of product j at time t by $r_t \in \{0, ..., 2^J - 1\}$ where we order the bundles such that $r_t = 0$ refers to no consumption (i.e., an empty bundle) and $r_t \in [1, J]$ refers to a bundle that contains only good $j = r$. The indirect utility individual i obtains from consuming j in market m at time t is given by

$$\overline{u}_{ijmt} = \alpha_j p_{jmt} + D'_{it}\pi_j + X'_{mt}\lambda_j + \nu_{ij} \tag{1}$$

where α_j, π_j and λ_j are (vectors of) parameters to be estimated. The choice of using a substance depends on the price, p_{jmt}, where the products are measured in different units (i.e., gram, pack, 6-pack) so α_j captures the willingness-to-pay for a unit of substance j.

Choices may also depend on individual-specific components. We include observed demographics which may influence use (i.e., gender, age, education, and race) in the vector D_{it}. The parameter vector π_j allows for demographics to influence use differently depending on the substance. Market-specific variables that impact choices are represented by X_{mt} and include the year in which the substance was purchased, the quality of marijuana available in the market, and the importance of police enforcement. Again, the parameter vector that captures the impact of the market variables, λ_j, varies by substance. This is important as there are different levels of enforcement for substance-related crimes (such as drunk driving) and these may change across years.

In addition, we allow for unobserved heterogeneity (ν_{ij}) that may influence a person's choice of using marijuana, smoking cigarettes or drinking. There are a couple of points to mention. First, we wish to determine the complementary nature among these goods in demand, and identification of this is confounded by the presence of unobserved heterogeneity. Unobserved heterogeneity is person-specific and likely to be correlated across products. To allow for this, we assume that the random effects are distributed tri-variate normal with a full (symmetric) covariance matrix. In addition, there may be unobserved terms that impact the price that are correlated with the quality of the product offered - i.e., price may be endogenous. We discuss identification including these issues in section 4.2.

Individuals obtain an indirect utility from consuming the goods in combination. Again, following Gentzkow (2007), the utility an individual obtains from consuming bundle r is

$$u_{irmt} = \sum_{\mathcal{S}_j \in r_t} \overline{u}_{ijmt} + \gamma_r + \epsilon_{irmt} \tag{2}$$

where the $\mathcal{S}_j \in r_t$ denotes the set of goods containing j that belong to bundle r. We assume that the idiosyncratic terms (ϵ_{irmt}) are distributed normally. The utility from the outside option of consuming none of the products is normalized to zero, because we cannot identify relative utility levels. The set of variables that result in consumption of bundle r given the parameters of the model is given by $A_{irmt} \equiv \{u_{irmt} \geq u_{ikmt} \; \forall r \neq k\}$.

There is one important caveat - individuals who do not have access to a source of marijuana cannot purchase it - that is, these individuals face a limited choice set in part because the substance is illegal. We control for this restriction following previous literature on limited choice sets (e.g., Sovinsky Goeree (2008) and Jacobi and Sovinsky (2016) in the context of marijuana access). Specifically, as in Jacobi and Sovinsky (2016), we observe data on whether an individual has access to marijuana (see section 2.1). Therefore, for individuals who have limited access, the set of bundle choices is limited to those without marijuana (denoted R^L), where the set of variables resulting in consumption of bundle $r \in R^L$ is given by : $A^L_{irmt} \equiv \{u_{irmt} \geq u_{ikmt} \; \forall r \neq k \text{ for } r, k \in R^L\}$.

We generate access based on individual reported access, so one concern is that unobservables are correlated across access and use (Jacobi and Sovinsky (2016)). We conducted robustness checks with a measure of access that is free from unobserved variation (but with less individual variation). We discuss this in section 7. The results of the robustness checks were not significantly different. The additional variation in our access variable aids in identification so we continue with self-reported access in the main specification.

3.1 Accounting for Past Substance Use

Augmenting the model to allow for persistence in consumption is straightforward, in theory. The main caveat is that we do not observe individual past substance use (denoted w^{t-1}_{ij}). However, we know something about the distribution of past choices of individuals (with similar demographics) from the PSID data. We use these to construct an empirical distribution of past behavior, given by $\hat{P}_j$, which we use to generate w_{ij}^{t-1}. We provide details of the data construction in 2.3 and estimation in section 4. We note that our methodology is in the spirit of Berry et al. (1995) which draws a "consumer" from an empirical distribution generated from another dataset.

We modify the indirect utility of individual i, $\overline{u}_{ijmt}$ given in equation (1) by adding a term that captures past use as follows:

$$w_{ij}^{t-1}\delta_j + p_{jmt}\alpha_j + D_{it}'\pi_j + X_{mt}'\lambda_j + \nu_{ij}, \quad w_{ij}^{t-1} \sim \hat{P}_j \tag{3}$$

where the parameter vector, δ_j, captures habit persistence that may differ across substances. The utility an individual receives from consuming a bundle is formed in an analogous manner to the static counterpart in equation (2). Given limited choice sets for some consumers, the probability i chooses to buy bundle r in period t is

$$s_{irmt} = \begin{cases} \int_{A_{irmt}} dF(\epsilon, \nu)\, d\hat{P}(w^{t-1}) & r \in R \\ \int_{A_{irmt}^L} dF(\epsilon, \nu)\, d\hat{P}(w^{t-1}) & r \in R^L \end{cases} \tag{4}$$

Our approach allows us to turn cross-sectional data into a "matched panel" to overcome data limitations by using information from more than one source to inform our estimates.

4 Econometric Methodology

We propose a multinomial probit framework for bundle choices that takes into account limited choice sets and persistence via the matched panel approach. Building on the Bayesian MNP literature (e.g., Albert and Chib (1993), McCulloch, Polson, and Rossi (2000), Kosuke and Van Dyk (2005), Loaiza-Maya and Nibbering (2022)), we proceed via the data augmentation approach to develop an efficient estimation strategy based on Gibbs-sampling that can be extended to estimate choice probabilities of bundle use under different policy settings, providing both point estimates and precision, as well as to obtain the bundle and substance participation elasticities. Similar to Jacobi and Sovinsky (2016), we combine the posterior estimation with the Bayesian predictive approach for the counterfactual policy analysis to obtain the distributions of substance use under various policy settings.

4.1 Joint Substance Use under Heterogeneous Choice Sets and Group-based Persistence

Let $y_{im} = r$ denote the observed categorical outcome variable for individual i (we omit the subscript t for ease of exposition), which reflects consumption of bundle r that is a combination of substances $j \in \mathcal{J}$, with $\mathcal{J} = \{1(\text{mar}), 2(\text{cig}), 3(\text{alc})\}$. One important caveat is that individuals who do not have access to marijuana cannot purchase it - that is these individuals face a limited choice set in part because the substance is illegal. We can account for this limited choice set because we observe marijuana access a_{im} in the data. Individuals who report access ($a_{im} = 1$) choose from the unrestricted set of consumption bundles

$$\mathcal{R} = \{1(\text{mar}), 2(\text{cig}), 3(\text{alc}), 4(\text{mar, cig}), 5(\text{mar, alc}), 6(\text{cig, alc}), 7(\text{mar, cig, alc})\}$$

where $r = 0$ is the base category of no consumption. Individuals without access to marijuana ($a_{itm} = 0$) can only choose from the limited set of bundles $\mathcal{R}^L = \{2, 3, 6\}$ that contains all bundles without marijuana and the outside option. The observed choice is modeled in terms of the latent bundle utilities $\tilde{u}_{irm}$ and observed access a_{itm} as

$$y_{im} = \begin{cases} r & \text{if } \tilde{u}_{irm} > max\{\tilde{u}_{irm}^{-r}, 0\} \\ 0 & \text{otherwise} \end{cases} \quad \text{where } r \in \begin{cases} \mathcal{R} & \text{if } a_{im} = 1 \\ \mathcal{R}^L & \text{if } a_{im} = 0 \end{cases} \tag{5}$$

where the latent utility is maximized over the relevant choice set given an individual's access (the latent utility of the outside good is normalized to zero). Corresponding to the identification assumption in Gentzkow (2007), the idiosyncratic bundle errors ϵ_{irmt} are assumed to be independently distributed and the unobserved tastes are jointly Normally distributed with covariance matrix $\mathbf{\Sigma}$ as

$$\boldsymbol{\nu}_i = (\nu_{1i}, \nu_{2i}, \nu_{3i})' \sim N(0, \mathbf{\Sigma}) \tag{6}$$

to capture in the correlation of bundle choices beyond the bundle effects.

Latent utility is defined in accordance with the economic model specification (2) as

$$\tilde{u}_{irm} = \sum_{j \in r} (\mathbf{Z}'_{im}\boldsymbol{\beta}_j + \nu_{ij}) + I[r > J]\gamma_r + \epsilon_{irm} = \mu_{irm} + \varepsilon_{irm} \tag{7}$$

where $\epsilon_{irm} \sim N(0,1)$. Further, we define $\mathbf{Z}_{im} = (p_{jm}, D_i, X_m)$ and $\boldsymbol{\beta}_j = (\alpha_j, \pi_j, \lambda_j)$ for the baseline specification that does not incorporate past use.

For the analysis with persistence we define $\mathbf{Z}_{im} = (I(j \in \boldsymbol{\omega}_k^{t-1}), p_{jm}, D_i, X_m)$ and $\boldsymbol{\beta}_j = (\delta_j, \alpha_j, \pi_j, \lambda_j)$ where $\boldsymbol{\omega}_k^{t-1}$ denotes group-specific past use generated via the matching approach. In particular, we consider a subset of discrete demographic variables $\tilde{D} \subset D$, that are observed for individuals both in the MTF and PSID data, to define the set of k demographic groups $\mathcal{G}(\tilde{D}) = \{G^k\}$. From the PSID data, we obtain the group-specific proportion of past use of substance j for subjects in group k, $\hat{\omega}_j^k$, and generate $\boldsymbol{\omega}_{jk}^{t-1} = \hat{\omega}_j^k \; \forall i \in G^k$. For the empirical analysis we specify $\tilde{D}$ to include three discrete variables (gender, race, aged under 18) which yields a total of eight groups.[25]

4.2 Identification

A common concern in discrete choice models of product choice is that there is some component not included in the utility function (such as quality) and so it appears in the error term. To the extent that prices are correlated with quality, this presents a potential endogeneity problem. We have three substances with associated prices - marijuana, cigarettes, and alcohol. First, we note that our framework differs from standard models of product choice because we are modeling the decision to consume a substance - not a brand of the substance. The prices we use for the substances are applicable to all brands of those substances. For marijuana, we use the (weighted) average price, where we include a control for quality (the amount of high-quality marijuana in the market) in the consumer's utility function. For cigarettes and alcohol, we use the tax rates which do not differ by brand. Finally, we control for region fixed effects to capture any location related unobserved heterogeneity that may influence tax rates as well as substance use.

Identification of the bundle effects is in the same spirit as that outlined in Gentzkow (2005, 2007). If we see that consumers use cigarettes and alcohol together one possibility is that the products are complements. However, another reason we could see these two products being consumed together could be that an individual has unobserved tastes for

[25] This implies that $f(\omega_{ij}^{t-1}) = \int \sum_{k=1}^{K} \hat{w}_k^{t-1} \, Pr(i \in G^k | \tilde{D}_i) \, d\tilde{D}_i$

alcohol and cigarettes that are correlated (i.e, the individual has a greater taste for having a high feeling). Gentzkow (2005, 2007) provides an extensive discussion of how to separately identify the unobserved covariance from substitution/complement parameter (γ). Exclusionary restrictions are one source of identification - something that impacts the utility of one product but not the other product. In our framework, prices of the products are relevant exclusionary restrictions. In addition, access to marijuana impacts the choice to purchase marijuana but not the utility from consuming alcohol or cigarettes so access also serves as an exclusionary restriction. Finally, the substitution patterns are also identified by variation in the choice set. This is particularly relevant for our case as the choice set is observed and varies depending on access.

Identification of the remaining parameters in the utility function are identified by variation in substance choices (from the set of products available) corresponding to variation in the observable attributes of the substance or the individual (such as prices or demographic characteristics). Finally, as previously discussed, we don't observe past choices of individuals in the MTF data. The parameters on past use are identified by variation in past choices of PSID individuals from a similar demographic cohort to the MTF individual corresponding to variation in the observable attributes of the substance.

4.3 Estimation Strategy

Assuming normally distributed bundle errors, the $R \times 1$ vector of latent bundle utilities $\tilde{\mathbf{u}}_{im} = (u_{i1m}, u_{i2m}, ..., u_{iRm})$ follows a normal distribution with $N(\boldsymbol{\mu}_{im}, I_7)$ for individuals with $a_{im} = 1$. For individuals with limited access, $a_{im} = 0$, we have a reduced vector of latent utilities from the limited choice set, $\tilde{\mathbf{u}}^L_{im} = (u_{i2m}, u_{i3m}, u_{i6m})$ with the distribution $N(\boldsymbol{\mu}^L_{im}, I_3)$. The likelihood of the observed bundle choices $\boldsymbol{y} = \{y_{im}\}$ given the observed access $\boldsymbol{a} = \{a_{im}\}$ and mean utility parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}_j, \gamma_j : j = 1, .., J\}$, $f(\boldsymbol{y}|\boldsymbol{a}, \boldsymbol{\theta}, \boldsymbol{\nu})$, can be expressed as

$$\prod_i \left\{ \int_{A_{irm}} N(\boldsymbol{\mu}_{im}, I_7)\, d\,\boldsymbol{u}_{im} \right\} I[a_{im} = 1] + \left\{ \int_{A^L_{irm}} N(\boldsymbol{\mu}^L_{im}, I_3)\, d\,\boldsymbol{u}^L_{im} \right\} I[a_{imt} = 0] \qquad (8)$$

where the integration region is given by $A_{irm} = \{\tilde{u}_{irm} > max\{0, \tilde{u}_{-rm\}} : r \in \mathcal{R}\}$ for individuals with an unlimited choice set and $A^L_{irm} = \{\tilde{u}_{irm} > max\{0, \tilde{u}_{-rm\}} : r \in \mathcal{R}^L\}$ for those facing a limited choice set. The two-part likelihood implies that we only learn about marijuana preferences from individuals with access to marijuana. Note that we do not estimate the parameters associated with having access (we treat those as predetermined) but we control for access when estimating the parameters of the choice set.

The above expression involves high-dimensional integrals that do not have an analytical solution. Proceeding with the data augmentation approach, inference on the model parameters conditional on the data $\boldsymbol{D} = \{\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{X}\}$, including the full set of covariates $\boldsymbol{X} = \{Z_{im}\}$, is directly based on the posterior distribution of the model parameters and the latent utilities $\tilde{\mathbf{u}} = \{\tilde{u}_{irm}\}$ which takes the form

$$\pi(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{u}}|\boldsymbol{D}) \propto \prod_i \Big(TN_{A_{irm}}(\boldsymbol{\mu}_{im}, I_7)I[a_{im} = 1] + TN_{A^L_{irm}}(\boldsymbol{\mu}^L_{imt}, I_3)I[a_{im} = 0]\Big)$$
$$\times\, N(\boldsymbol{\theta}|\mathbf{b}_0, \mathbf{B}_0)\, N(\boldsymbol{\nu}|0, \boldsymbol{\Sigma})\, IW(\boldsymbol{\Sigma}|v_0, R_0) \tag{9}$$

where the augmented likelihood becomes a product over truncated normal distributions, as before with two components depending on the observed access.

Under the standard assumptions of Normal prior distributions for the vector of regression coefficients, $\boldsymbol{\theta}$, and inverse Wishart prior for $\boldsymbol{\Sigma}$, draws from the posterior distribution of the model parameters can be simulated via an efficient Gibbs algorithm (see Appendix B.1). Point and interval estimates of the model parameters are computed directly from 10,000 draws from the described MCMC algorithm (after the initial burn-in phase). In the results section we report the posterior mean estimates of the model parameters alongside the posterior standard deviations and posterior credibility intervals.

4.4 Counterfactual Policy Inference

Key parameters for the policy analysis are the participation price (tax) elasticities of demand and the probability of substance use under different policy scenarios. Both can be obtained within the Bayesian predictive framework which incorporates the information from the posterior distribution of the model parameters.

Let $y^p_{im} = r$ denote the predicted bundle choice of individual i under some policy setting $\mathcal{P} = (\{a^{\mathcal{P}}_{im}, X^{\mathcal{P}}_{im}, p^{\mathcal{P}}_{jm}\})$ that is characterized by a specific combination of individual access that defines the choice set, as well as market characteristics (police enforcement, marijuana quality) and the vector of prices/taxes in the utility equation. The probability of use of substance S_j^p under a policy scenario $\mathcal{P}$ is then the proportion of all subjects whose predicted bundle choice contains substance j. Its distribution can be expressed as

$$f(S_j^p|\mathcal{P}) = \int g(S_j^p|\boldsymbol{y}^p) \prod_{i=1}^{n} f(y_{im}^p, \tilde{\boldsymbol{u}}_{im}|\mathcal{P}, \boldsymbol{\Theta})\, \pi(\boldsymbol{\Theta}, \tilde{\boldsymbol{u}}|\boldsymbol{D})\, d\boldsymbol{\Theta}\, d\tilde{\boldsymbol{u}} \tag{10}$$

where $g(S_r^p|\boldsymbol{y}^p) = \frac{1}{n}\sum_{i:j\in r}^{n} I[y_{im}^p = r]$ with the predicted bundle choices and latent bundle utilities generated from the augmented likelihood $f(y_{im}^p, \tilde{\boldsymbol{u}}_{im}|\mathcal{P}, \boldsymbol{\Theta})$ conditional on the specific policy setting and the model parameters vector $\boldsymbol{\Theta} = \{\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\Sigma}\}$ coming from the posterior distribution. Similarly, we compute the bundle use probabilities as a proportion of all subjects whose predicted bundle choice is bundle r with the corresponding distribution $f(S_r^p|\mathcal{P})$ obtained via expression (10) using $g(S_r^p|\boldsymbol{y}^p) = \frac{1}{n}\sum_{i}^{n} I[y_{im}^p = r]$.

Under policy scenarios with easing where police enforcement is removed, the value of the variable is set to zero for all individuals. In policy settings with full marijuana access, we set $a_{im}^{\mathcal{P}} = 1$ for all individuals. In scenarios with "high quality" marijuana, the variable is set to one for individuals when computing the latent utilities. Changes in taxation are implemented relative to the 2013 taxes and prices.

In specifications without persistence, the counterfactual analysis does not contain a lagged component. Hence, bundle and substance use probabilities remain constant after the initial impact of the policy change. For the analysis with persistence, we estimate a sequence of predicted probability of use $\{f(S_{jt}^p|\mathcal{P}, w_{kj}^{t-1}\}$ from $t = T, T+1, ..., T+5$, where $t = T+1$ denotes the first period after the policy change. For the first prediction period, $t = T$, w_{kj}^{t-1} is based from the empirical distribution based on the matched sample approach. From $t = T+1$ onward, past use is based on the empirical distribution from the predicted use in $t-1$ based on the set of demographic groups $\mathcal{G}(\tilde{D})$ as defined in Section 4.1. In the implementation, we designate year 2013 as the base year $T = 2013$, and adjust all observations from 2004 to 2013 to this base year by applying the 2013 marijuana prices, cigarette and alcohol taxes, and the year's fixed effects. We then use the past use constructed based on the empirical distribution of year $T-1 = 2012$ to predict substance use for the year 2013. This method is internally consistent. Moreover, our predicted use for the year 2013 aligns with the observed use in the MTF, lending validity to our approach.

The integral in (10) is solved via the method of decomposition using the MCMC parameter draws to generate the latent bundle utilities and bundle choice given those draws and to compute the probability of use for each substance. This is implemented via an additional set of draws at the end of each iteration of the Gibbs sampler. To obtain the distribution of the participation price elasticities, we combine the Bayesian predictive approach for substance use estimation with two-sided numerical integration to assess the derivatives and obtain the distributions of the corresponding participation elasticities. Details are provided in Appendix B.2.

5 Multi-Substance Use Results

We begin with the results for our model without persistence, presented in Table 5. First, we discuss the parameter estimates for the product characteristics and demographics. Next, we examine what the results imply for the complementary nature among the substances. Then we highlight the importance of controlling for restrictions in access to marijuana and persistence in use.

5.1 Price, Quality, Enforcement, and Demographics

The first columns of Table 5 provide the results for our main baseline specification - where we account for the limited choice set due to restricted access to marijuana. Not surprisingly, the results indicate that prices matter in consumption decisions across all products. With respect to marijuana usage, the marginal valuation of quality is positive (0.154), whereas the marginal utility of marijuana consumption is negatively related to the concern of being arrested (police enforcement) (-0.282). Nearly all of the density of the (posterior) distribution of the quality coefficient is positive (96%), while the contrary is true of the enforcement density (96% of the density is negative). We conduct counterfactual experiments in Section 6 that quantifies the importance price, quality, and enforcement play in consumption decisions.

The results also show that demographic characteristics matter in choice of substances but not necessarily in the same direction for all products. For example, individuals who have

Table 5: Multi-substance Use Estimates

	Limited Access (Limited Choice Set)			No Choice Set Restrictions (Full Choice Set)		
	Marijuana	**Cigarettes**	**Alcohol**	**Marijuana**	**Cigarettes**	**Alcohol**
Product and Market Characteristics						
Price	-0.014*	-0.088***	-0.005***	-0.049***	-0.072***	-0.014***
	(0.007)	(0.008)	(0.001)	(0.010)	(0.012)	(0.001)
High Quality Proportion	0.154			0.624***		
	(0.117)			(0.165)		
Police Enforcement	-0.282**			-0.570***		
	(0.112)			(0.157)		
Demographics						
White	-0.204***	0.329***	0.191***	-0.029	0.579***	0.430***
	(0.013)	(0.010)	(0.010)	(0.037)	(0.036)	(0.038)
Male	0.246***	0.016*	-0.012	0.355***	0.167***	0.123***
	(0.012)	(0.009)	(0.009)	(0.017)	(0.024)	(0.016)
Aged 18 and over	-0.047***	0.092***	0.025***	-0.032**	0.138***	0.055***
	(0.012)	(0.009)	(0.009)	(0.014)	(0.015)	(0.015)
Max Parental Edu: HS	-0.074***	-0.117***	0.083***	-0.096***	-0.178***	0.084***
	(0.018)	(0.013)	(0.014)	(0.018)	(0.021)	(0.024)
Max Parental Edu: College	-0.058***	-0.237***	0.114***	-0.142***	-0.371***	0.085***
	(0.016)	(0.012)	(0.012)	(0.017)	(0.022)	(0.023)
Bundle Effects						
With Cigarettes	0.924***			0.286***		
	(0.013)			(0.039)		
With Alcohol	2.001***			1.567***		
	(0.016)			(0.237)		
Cigarettes and Alcohol	0.931***			0.672***		
	(0.012)			(0.106)		
Unobserved Heterogeneity						
Marijuana	1.000[a]	0.050***	-0.693***	1.000[a]	1.003***	-0.138
Cigarettes	0.050***	0.005***	-0.035***	1.003***	1.28***	0.621***
Alcohol	-0.693***	-0.035***	0.483***	-0.138	0.621***	2.308***

Notes: Mean coefficients, standard deviations in parenthesis. Year fixed effects are included. Region fixed effects are included for bundles with marijuana. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The significance levels are based on the posterior credibility intervals. [a] Normalized to 1.

higher educated parents are more likely to drink alcohol (0.114) but less likely to smoke cigarettes (-0.237) or use marijuana (-0.058). Whites are less likely to consume marijuana (-0.204) but are more likely to drink (0.191) and smoke cigarettes (0.329) relative to non-whites. Gender also exhibits differential use patterns depending on the substance, with males more likely to consume marijuana (0.246) and smoke (0.016) relative to females. Youth aged 18 and older are less likely to consume marijuana (-0.047) but more likely to drink alcohol (0.025) and smoke cigarettes (0.092), relative to those under the age of 18.

5.2 Sin Goods: Complements or Substitutes?

The bottom panel of Table 5 presents the estimates for the bundle effects and unobserved heterogeneity in substance use. The results show that individuals obtain additional positive utility from using marijuana with cigarettes (0.924), marijuana with alcohol (2.001), and cigarettes with alcohol (0.931). In addition, unobserved tastes are correlated across goods. On average, individuals' unobserved tastes for consuming marijuana and cigarettes are positively correlated (0.050), while they are negatively correlated for marijuana and alcohol (-0.693). Finally, unobserved tastes are negatively correlated for alcohol and cigarettes (-0.035). We use these estimates to compute the price elasticities, which allows us to determine the complementary nature of the products.

Table 6: Participation Substance Elasticities

	Limited Choice Set Substance Use			**Full Choice Set** Substance Use		
	Marijuana	**Cigarettes**	**Alcohol**	**Marijuana**	**Cigarettes**	**Alcohol**
Marijuana	-0.185	-0.092	-0.069	-0.472	-0.049	-0.083
	(0.097)	(0.048)	(0.036)	(0.102)	(0.020)	(0.019)
Cigarettes	-0.081	-0.233	-0.056	-0.010	-0.108	-0.008
	(0.009)	(0.022)	(0.006)	(0.005)	(0.020)	(0.003)
Alcohol	-0.055	-0.052	-0.068	-0.049	-0.025	-0.083
	(0.009)	(0.009)	(0.010)	(0.009)	(0.007)	(0.008)

Notes: Cell entries (i, j), where i indexes row and j indexes column, give the mean (standard deviation in parentheses) percentage change in the probability of using j substance in response to a percentage increase in the price of substance i.

Table 6 presents the estimates for own and cross-price (tax) participation elasticities for the substances.[26] The entries represent the percentage change in the probability of using the column substance in response to a one percentage change in the price of the row substance. The first columns contain the elasticities implied by our preferred specification that controls for restricted access. We find that a 10% increase in the price of marijuana results in a decline in the probability of marijuana use of 1.85% We find that youth cigarette use is sensitive to tax, where the probability of use declines by 2.3% with a 10% tax increase. Finally, participation in alcohol shows the least sensitivity to a tax increase (declines by 0.7% with

[26] We compute the elasticities by changing the price by 1% and predicting the change in the probability of use. We provide details in Appendix B.2.

a 10% tax increase). We note that these own-price/tax elasticities are consistent with the range of participation elasticities from the literature. In particular, marijuana participation elasticities assuming full choice sets range from -0.48 (Pacula et al. (2001)) to -0.21 (Zhao and Harris (2004)), while Jacobi and Sovinsky (2016), who control for access selection, estimate a marijuana price elasticity of -0.20. DeCicca, Kenkel, and Mathios (2002), Carpenter and Cook (2008) and Sen and Wirjanto (2010) find cigarette participation elasticities ranging from -0.14 to -0.72. And, using data from an older cohort, Dave and Saffer (2008) find an alcohol participation elasticity of -0.04.

Finally, we can answer the main question of the paper - are sin goods complements or substitutes? The results point to the former - on average the products are complements. This can be seen by the cross-price elasticities given on the off-diagonals, which are negative for all substance interactions. They show that the probability of using marijuana declines by 0.81% with a 10% increase in the tax on cigarettes and by 0.55% with a 10% increase in alcohol tax. Likewise, cigarette use and alcohol use decline as the prices of other substances increase. The complementary nature of the products implies that the government has two additional policy variables to control marijuana use - the tax rate of alcohol and the tax rate of cigarettes. We explore the implications in section 6.

For comparison, we estimated a multi-variate model to examine the resulting implications for substitution (under no access restrictions to be comparable to prior literature). This could be worthwhile as a multi-variate model is simpler to estimate, and it allows for correlation across choices of products through unobserved correlation and via shared regressors in each choice equation. (However, the choice between substances is otherwise independent.) We estimated two specifications, one where the prices of all products enter all choice equations and a second where all prices and marijuana quality and police enforcement enter all equations. The estimated elasticities show no significant relationship between marijuana and the other substances. As Table 5 shows, in our framework, the bundle effect estimates are positive but the unobserved heterogeneity is positive in some cases and negative in others. The multi-variate framework has no way to disentangle these two effects.

5.3 Role of Restricted Access and Persistence in Use

Limited Choice Sets One contribution of our paper is to control for restricted access in marijuana use due to the illegal nature of the substance. The final columns of Table 5 provide the estimates if we remove this access restriction. Individuals appear to be more price sensitive to marijuana (-0.049 relative to -0.014), to value the quality more (0.624 relative to 0.154), and to be more concerned with being arrested (-0.570 relative to -0.282). We also see differences in the impact of demographics on the marginal utility of consumption (in magnitude and significance, although not in direction). Furthermore, the distribution of the price coefficient is shifted more to the left and more dispersed under the full choice set specification. This implies we would be likely to overestimate the price impact under a full choice set assumption.

The differences in parameter estimates for the full choice set translate into different estimated elasticities, as shown in the latter columns of Table 6. Specifically, marijuana own price elasticities are more than double under the full choice set assumption (-0.472). However, the cross-price elasticities are negative, implying that the choice set restriction does not change the finding of the complementary nature of the products although the full access model does yield substantially different cross-price sensitivities relative to the limited choice set model. In particular, full choice set marijuana cross-price elasticities are less sensitive to changes in prices of alternative products, which would generate incorrect conclusions regarding the effectiveness of taxing alternative products on use. In summary, we find it is important to control for access in use to evaluate the impact of taxing polices, but less crucial for determination of the direction of substitution.

Persistence in Use The final addition is to allow for persistence in use given the potential addictive nature of the substances. For ease of exposition, we restrict our discussion to the limited access specification and focus on the lagged use parameter estimates and the implied price elasticities. For reference, we provide estimates in Table A1 in Appendix C. Recall that our measure of past use is based on the past use for a similar demographic cohort. We find that past use is an important determinant of current use. In particular, the estimates of lagged use are 0.648 for marijuana, 0.355 for cigarettes, and 0.116 for alcohol.[27]

The literature generates a broad range of estimates of the impact of past substance use on current use. However, none do so within a three-substance bundle model. Our estimates fall within these ranges for all substances.[28]

The remaining parameter estimates do not differ substantially from those in the static model (Table 5) neither in terms of characteristics nor bundle effects. As a result, the implied price elasticities from the model with persistence show very slight changes in the substitution patterns relative to the static framework. The cross-price elasticities continue to indicate all substances are complements. The own-price elasticities are also not significantly different than those from the static model.[29]

In summary, there are two main findings that emerge from the results. First, we find evidence for complementary effects between all substances. Second, controlling for access shows participation is less sensitive to price changes. These findings have implications for policy analysis, which we examine in the next section.

6 Counterfactual Policy Analysis

Ultimately this process yields insight into the impact on substance use arising from changes in regulation of potentially different sin goods. We consider the impact of a number of scenarios (and combinations) implemented in a counterfactual world where marijuana is legal across the U.S. We begin by examining different taxing strategies under the status quo. Next, we consider easing of legality and access restrictions, as well as changes in the prices and quality of marijuana. Then we examine how tax rates can be used to control substance use. Under each of these settings, we predict the impact of the policies in the longer term. Finally, we examine what set of tax rates and legality restrictions could be used in combination to curb use among underage youth. For all counterfactuals, the status quo

[27] We note that the coefficient for lagged alcohol use does not lie in the 10% Bayesian credibility interval.

[28] In particular, lagged use of marijuana ranges from 0.09 to 1.77; cigarettes 0.10 to 1.12; alcohol -0.03 to 1.81. See Deza (2015) for marijuana and alcohol. Becker et al. (1994); Chaloupka (1991); Darden (2017); Gilleskie and Strumpf (2005); Gordon and Sun (2015); Hai and Heckman (2022); Labeaga (1999); Ma (2017) are comparable studies for cigarettes; Goel and Morey (1995); Heien and Durham (1991); Pierani and Tiezzi (2009) for alcohol and cigarettes, and Grossman et al. (1998); Waters and Sloan (1995) for alcohol alone.

[29] They are -0.186, -0.235, and -0.068, for marijuana, cigarettes, and alcohol, respectively.

is evaluated at prices in the last period of our data (2013). We take each of these in turn.

Tax Policies Marijuana sales in the US are regulated at the state level, where the variation across states is large. For example, as of April 2023, recreational marijuana excise tax rates ranged from 6% (Missouri) to 37% (Washington).[30] We will start with the population-weighted average marijuana tax rate, which is around 20%.

Figure 2: Impact of Tax Changes on Substance Use

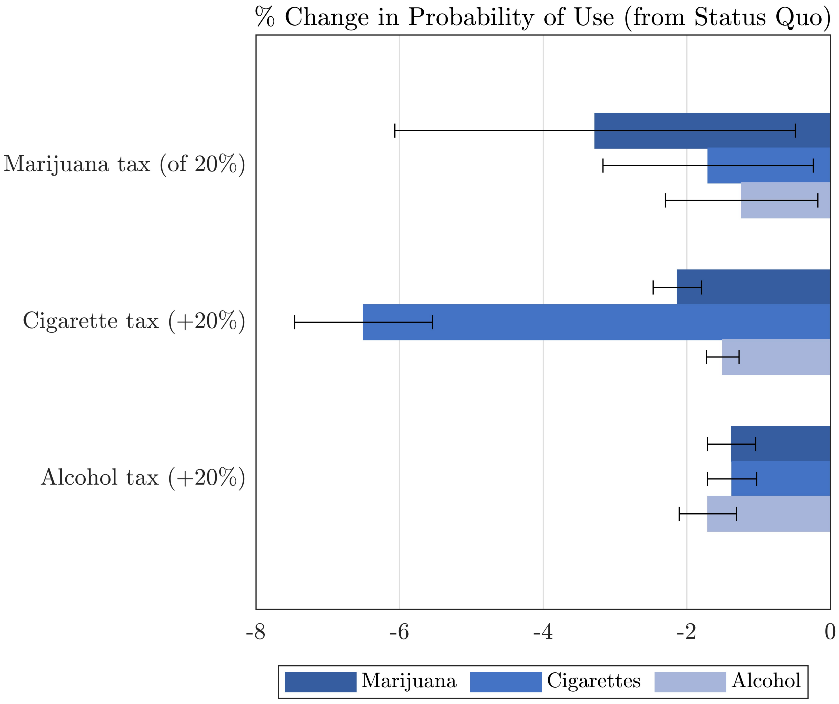


Figure 2 presents the changes in the probability of substance use under alternative tax scenarios relative to pre-legalization levels of use (status quo). The darkest bar shows the change in the prevalence of marijuana use; the lightest the change in alcohol use; and the middle the change in cigarette use. The black bands show the 90% Bayesian credible interval of the predictions. Note that, in this set of counterfactuals, we assume that nothing else in the environment has changed - in particular marijuana is still illegal. Hence, we interpret the increase in the "tax" on marijuana as an increase in the price of marijuana, which is

[30] This is calculated as percentage of the retail price. Source:https://taxfoundation.org/data/all/state/state-recreational-marijuana-taxes-2023/)

useful as a benchmark for (later) counterfactuals where use is legal and taxed.

Figure 2 shows that increasing taxes results in less consumption of the substance, in line with the own-price elasticities. Furthermore, we see that consumption of other substances declines, reflecting the complementary nature across substance use. For example, the middle set of bars shows that a 20% increase in the tax on cigarettes results in a decline of 6.5% in cigarette consumption, as well as a decline of 2.1% in marijuana use. Likewise, a 20% increase in the tax on alcohol would reduce marijuana use (by 1.4%).

Marijuana Legalization Policies Under legalization, marijuana use would continue to be illegal for youth under the age of 21. However, easing of restrictions may impact youth use for a few reasons. First, it may become easier to access, as it will be available legally and not only via dealers. Hence, youth under 21 will face similar barriers to use as they face with alcohol. Second, use may change if police reduce their enforcement or if the youth are less concerned about enforcement (perhaps due to legality "normalizing" marijuana use). Finally, use may change if the characteristics of the product change, such as prices or quality.[31] Indeed, (non-edible) marijuana sold in legal dispensaries typically contains THC levels well above 15%, which we classify as high quality.[32] In order to examine the impact of the easing of legalization, we use the parameter estimates to conduct counterfactual policy experiments where we decompose changes in use due to: a reduction in police enforcement, reduction in access restrictions, and changes in prices and quality levels.

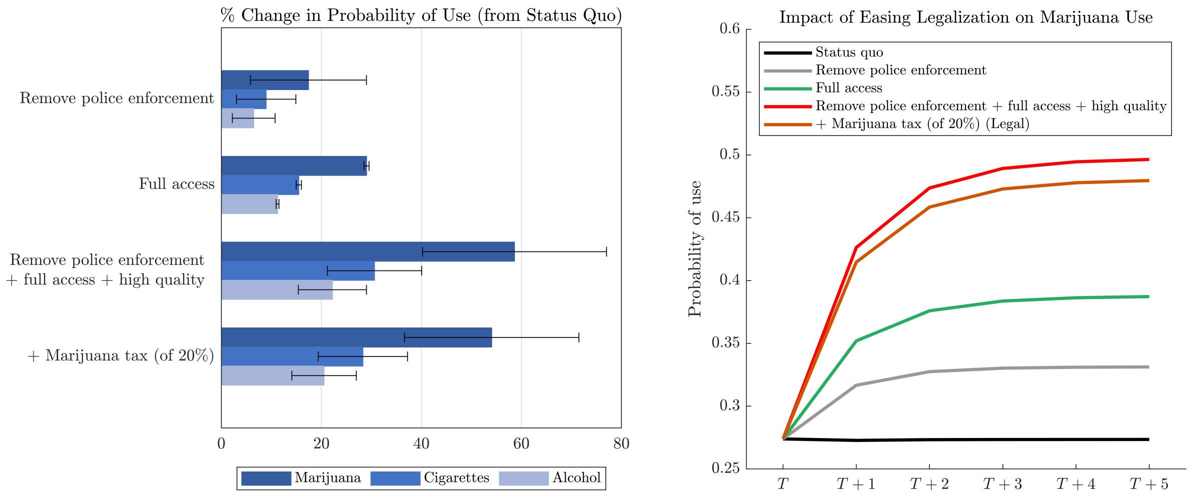
Figure 3a[33] presents the changes in the probability of substance use under alternative

[31] While we are able to incorporate changes in the quality of marijuana post-legalization, we cannot control for new product entry. However, to the extent that consumers value quality, and taxing policies on the new products are similarly implemented, our findings on joint use and how taxing policies on one product impact use of the other product are germane.

[32] Laws differ across states regarding the maximum amount of THC in a product. For example, the maximum THC allowed in flowers in Montana is 35% (https://mtrevenue.gov/cannabis/faqs/ Accessed Dec 5 2022). In 2017 the most popular strains found in dispensaries in Colorado had a THC-range of 17–28% (www.leafly.com). Grigolon, Plúas-López, Remmy, and Sovinsky (2022) examine the impact of quality changes as well as other product characteristics post-legalization on black market sales in Canada.

[33] The difference in the credible interval widths between full access and other policy tools stems from the different natures of the variables in the model. The effects of other policy tools are tied to specific coefficients, like policy enforcement or price effects, leading to wider intervals due to uncertainties in these estimates. Conversely, marijuana access directly influences individuals' choice sets and utility-maximizing

Figure 3: Deconstructing Legalization Policy Tools



(a) (b)

legalization scenarios relative to pre-legalization levels (status quo). The first set shows that removing police enforcement (which we implement by setting the variable to zero and re-calculating use) results in an increase in use of all substances: a 17% increase for marijuana, a 9% increase for cigarettes, and a 6.5% increase for alcohol. Removing access barriers alone has an even larger impact, with increases of 29%, 15%, and 11% for the three substances, respectively. Finally, removing access barriers and police enforcement, in combination with making the product high quality, indicates marijuana use would increase by nearly 60% from pre-legalization levels. Given that legalization is accompanied by a tax on sales, the final set of bars presents the most realistic view of the post-legalization use (hereafter, the "legal environment"). The results indicate that marijuana use would increase by around 54% after legalization. Furthermore, use of cigarettes and alcohol would increase by 28% and 20%, respectively. Our results in are line with observed use of sin goods in states where marijuana has been legalized. That is, as states have legalized marijuana, alcohol use and cigarette use has increased in those states.[34] In summary, our findings show that both restricting access and enforcing laws related to illegal use, are policy tools that can be effective in curbing use

decisions. This direct impact, when aggregated at the market level, results in a more certain effect for full access compared to other policy tools.

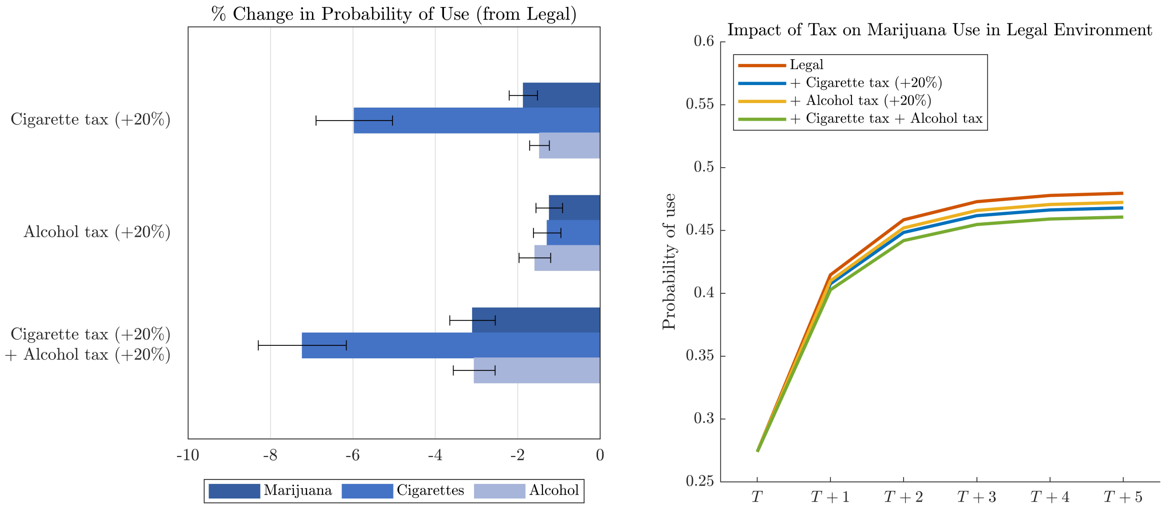
[34] See Macha, Abouk, and Drake (2022), Melillo (2022) and Kissi (2021). We provide more substantial evidence in section 7 that our results are consistent with observed use in areas where use has been legalized.

among the youth.

Figure 3b[35] shows how use would evolve over time (where we also allow the individuals to age appropriately). As the figure shows, the impact of legalization grows over time, but at a slower rate than the initial legalization period. This reflects the persistent nature of use.

Tax Policies under Legalization One way to control use is to increase marijuana tax directly, however, as the tax rate increases, individuals may move to the black market. Therefore, it is useful to understand how much tax on alternative substances can help to control use. Figure 4 shows the impact of taxing alternative substances simultaneously on marijuana use. As the lower set of bars in Figure 4a show, increasing the tax rate on alcohol and cigarettes by 20% in combination with taxing marijuana at 20%, results in a decline of marijuana use by 3% relative to a legal scenario where only marijuana is taxed. Furthermore, the impact of initial tax increases remains over time, as shown in Figure 4b.

Figure 4: Impact of Multi-Substance Tax in Legal World

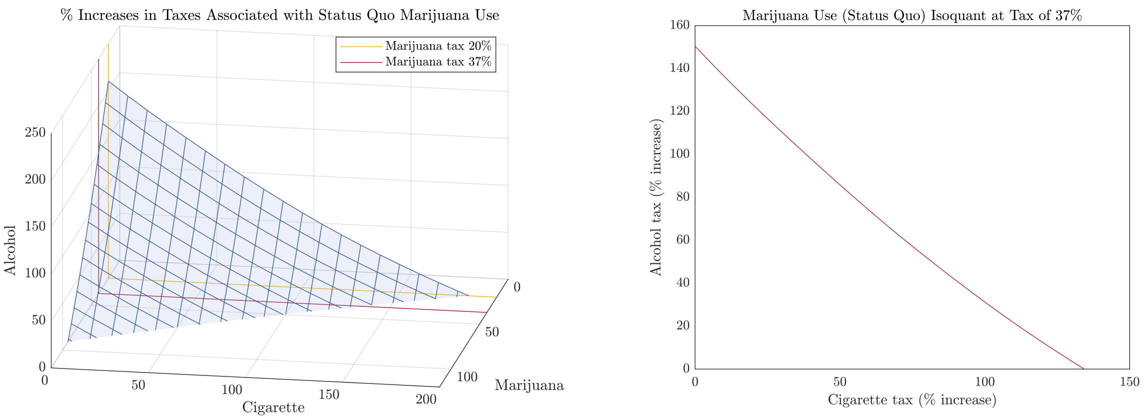


(a) (b)

[35] As discussed in Section 4.4, the probability of use in the year $T = 2013$, using past use constructed based on the empirical distribution from the year $T - 1 = 2012$, is consistent with the observed usage in the MTF data. This consistency lends validity to our approach.

Policies to Limit Youth Use A tax increase of 20% is rather arbitrary. Also, we do not have to apply tax and non-tax policy tools in isolation. While it is difficult to control access to marijuana, it is likely more feasible to continue to enforce illegal use among underaged youth. To that end, we examine what combinations of taxing strategies together with continued pre-legalization levels of police enforcement can be used to curb marijuana use - in particular, to return it to pre-legalization levels of use.

Figure 5: Tax Rates to Control Marijuana Use



(a) (b)

Figure 5 presents the results. The plane, in the left panel figure 5a, shows all combinations of tax increases needed to return use to pre-legalization levels assuming police continue to enforce illegal use. The yellow (red) line denotes the average marijuana (maximum) tax rates on marijuana post-legalization. Figure 5b shows the slice at a marijuana tax rate of 37%. The findings show that increases in tax on alcohol and cigarettes with maximums of 150 and 130 respectively would reduce use to pre-legalization levels.

The tax increases at first glance may seem large. However the magnitudes are not out of the observed ranges in the last decades - which have seen increases of both cigarette and alcohol state taxes of over 200%. For example, Illinois increased the beer tax by 171% in 2000; Alaska by 206% in 2003; and Nevada by 78% in 2004. More recently, Delaware increased the beer tax by 63% (2018). Significant increases in cigarette taxes are more recent. In 2017, California and West Virginia increased the cigarette taxes by 230% and 118% respectively, whereas Oregon increased it by 150% in 2021.

A case in point, in 2017 marijuana was legalised in Nevada. In addition, in 2017 Nevada increased cigarette taxes by 125%. Our results imply that if Nevada would have increased cigarette taxes simultaneously by only 8% (together with an increase of the marijuana tax to 37%) youth use would have been close to pre-legalization levels. [36]

One caveat, our model is presented in terms of taxing policies to yield clean policy implications. However, retail prices obviously matter as well (as taxes), and these are determined strategically. In particular, firms may react to tax changes by passing through all (or a portion of) the tax changes to consumer via changes in the retail price. To the extent that firms absorb (over pass through) tax changes (instead of passing them on exactly to consumers), our findings of tax implications for use will be overstated (understated).[37]

In summary, we find that there are reasonable levels of taxes increases on sin goods when, combined with a higher tax on marijuana and continued police enforcement, can reduce underage use to close to pre-legalization levels. Surprisingly, marijuana legalization has not been accompanied by changes in both alcohol and cigarette tax rates in any state, and only one state has increased one (Nevada). This suggests that vital policy tools to control youth use are being overlooked. This is particularly important given that our findings show that the earlier these tax increases are implemented the better the long-run outcome.

7 Goodness-of-Fit and Robustness

In this section, we first conduct a goodness-of-fit test to measure how well our model predicts use in a country where marijuana is currently legal, Canada. Second, given the importance of the restriction in access, we provide robustness checks under a variety of measures of access, including a measure that is free from potential endogeneity of access and use.

Canada was the second nation (after Uruguay) to legalize marijuana. Given that marijuana was legalized nationwide and that Canadian demographics are similar to those of the

[36] Sources: *Alcohol Policy Information System* (2023); *Delaware Department of Finance* (2023); Kenkel (2005); *Tax Foundation* (2023) In 2017 Nevada the marijuana tax combined rate was around 25%.

[37] There is evidence that alcohol taxes are exactly passed through to consumers, while results on cigarette pass-through are more mixed. See Kenkel (2005), Nelson and Moran (2020), Harding, Leibtag, and Lovenheim (2012), and DeCicca, Kenkel, and Liu (2013).

US, Canadian outcomes under legalization provide an opportunity to gauge the validity our predictions. In June 2018, the Canadian Senate passed a bill to legalize marijuana, and by 2019 it was possible to buy legally. Surveys reported by the Canadian Statistical Office show that prevalence rates among youth increased post legalization (e.g., StatCan (2020) and Zuckermann et al. (2021)). The largest increase (69%) was seen among youth who were in high school: who used at 15.9% in 2016-2017 (before it was legal) and at 27% in 2019 (when they were able to buy it legally). Among a wider age range of 12-24 years, use increased from 10.7% to 16.3% after legalization - an increase of 52%. Our counterfactuals predict that youth use would increase by 60% post legalization - which lies in the observed range in Canada. These findings provide anecdotal evidence that our model provides reasonable estimates of the impact of legalization.[38]

In Section 2.1 we discuss our measure of access to marijuana. We conducted robustness checks using both less and more restrictive definitions of access. Specifically, individuals who report having a fairly difficult time finding marijuana or who see drug dealers less than once a month in their neighborhood are defined as having access for our less restrictive specification (whereas our main specification of access assumes these individuals do not have access). This less restrictive definition implies about 84% of the sample has access to marijuana (relative to 77% for our main specification). The more restrictive access definition assumes instead that only individuals who find it very easy to get marijuana or see drug dealers in their neighborhood daily or almost daily have access to marijuana. This definition implies a lower probability of having access to marijuana (about 58%). The estimates and implied elasticities under these two alternative access specifications are consistent with those of our main specification, and they also indicate sin goods are complements in demand.

Furthermore, we were concerned that there may be unobserved correlation between access and use that we do not account for in our exogenous access framework that could arise as individuals self-report access and use. To address this potential issue we estimated an additional specification in which we first estimated a hedonic access variable as a function of demographics. We then include the predicted access variable as our measure of access. This

[38] The average price increase between 2018 and 2019 was approximately 16%.

definition of access does not include unobserved heterogeneity, and therefore is not potentially correlated with individual unobserved heterogeneity that may impact choices. The results are not significantly different from our main specification, but our main specification has more variation in access which is useful to aid in identification, motivating our decision to proceed with our main specification.

8 Conclusions

As illicit substances move into legal product space, substitution patterns with legal products become more salient. We focus on a young person's choice to consume marijuana, alcohol or cigarettes (and combinations) within a framework of multi-substance use. We estimate a structural model of multi-product use of illegal and legal substances that incorporates joint use, controls for limited access to illicit products, and allows for persistence in use.

The results suggest that marijuana, cigarettes, and alcohol are complements in consumption. Our results indicate it is important to control for access in use to evaluate the impact of taxing policies, but less crucial for the determination of the direction of substitution. Marijuana legalization will impact use of all substances, where use of marijuana among youth would increase from 25% to 38% under legalization. We show that the increase in use is due to decreased police enforcement as well as decreased access barriers.

Our counterfactual results show that both restricting access and enforcing laws related to illegal use, are policy tools that are effective in curbing use among the youth. We find that there are reasonable levels of tax increases on sin goods when, combined with a higher tax on marijuana and continued police enforcement, reduce underage use to close to pre-legalization levels. For example, Nevada increased alcohol taxes post-legalization - had they also increased cigarette taxes by 8% (coupled with a 8% higher marijuana tax), youth use would have been close to pre-legalization levels.

Surprisingly, marijuana legalization has not been accompanied by changes in both alcohol and cigarette tax rates in any state. This suggests that vital policy tools to control youth use are being overlooked. This is particularly important given that our findings show that the earlier these tax increases are implemented the better the long-run outcome.

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, *88*(422), 669–679.

Alcohol Policy Information System. (2023). Retrieved September 2023, from `https://alcoholpolicy.niaaa.nih.gov/`.

Arnabal Rocca, L. R. (2022). Legalizing harmful drugs: government participation and optimal policies. *The BE Journal of Economic Analysis & Policy*, *23*(1), 113–164.

Auriol, E., Mesnard, A., & Perrault, T. (2023). Weeding out the dealers? the economics of cannabis legalization. *Journal of Economic Behavior & Organization*, *216*, 62–101.

Becker, G., Grossman, M., & Murphy, K. (1994). An empirical analysis of cigarette addiction. *American Economic Review*, *84*(3), 396–418.

Becker, G., & Murphy, K. (1988). A theory of rational addiction. *Journal of political Economy*, *96*(4), 675–700.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 841–890.

Cameron, L., & Williams, J. (2001). Cannabis, alcohol and cigarettes: substitutes or complements? *Economic Record*, *77*(236), 19–34.

Carpenter, C., & Cook, P. J. (2008). Cigarette taxes and youth smoking: new evidence from national, state, and local youth risk behavior surveys. *Journal of health economics*, *27*(2), 287–299.

Chaloupka, F. (1991). Rational addictive behavior and cigarette smoking. *Journal of Political Economy*, *99*(4), 722–742.

Ching, A., Erdem, T., & Keane, M. (2009). The price consideration model of brand choice. *Journal of Applied Econometrics*, *24*(3), 393–420.

Ching, A., & Hayashi, F. (2010). Payment card rewards programs and consumer payment choice. *Journal of Banking & Finance*, *34*(8), 1773–1787.

Clements, K. W., Lan, Y., & Zhao, X. (2010). The demand for marijuana, tobacco and alcohol: inter-commodity interactions with uncertainty. *Empirical Economics*, *39*, 203–239.

Clerides, S., & Courty, P. (2017). Sales, quantity surcharge, and consumer inattention. *Review of Economics and Statistics*, *99*(2), 357–370.

Darden, M. (2017). Smoking, expectations, and health: a dynamic stochastic model of lifetime smoking behavior. *Journal of Political Economy*, *125*(5), 1465–1522.

Dave, D., & Saffer, H. (2008). Alcohol demand and risk preference. *Journal of Economic Psychology*, *29*(6), 810–831.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*(1-2), 109–126.

DeCicca, P., Kenkel, D., & Liu, F. (2013, 05). Who Pays Cigarette Taxes? The Impact of Consumer Price Search. *The Review of Economics and Statistics*, *95*(2), 516-529.

DeCicca, P., Kenkel, D., & Mathios, A. (2002). Putting out the fires: will higher taxes reduce the onset of youth smoking? *Journal of Political Economy*, *110*(1), 144–169.

Delaware Department of Finance. (2023). Retrieved September 2023, from `https://financefiles.delaware.gov/docs/abc.pdf.`

Deza, M. (2015). Is there a stepping stone effect in drug use? separating state dependence from unobserved heterogeneity within and between illicit drugs. *Journal of Econometrics*, *184*(1), 193–207.

Dills, A. K., Goffard, S., & Miron, J. (2016). Dose of reality: The effect of state marijuana legalizations. *Cato Institute Policy Analysis*(799).

Dills, A. K., Goffard, S., & Miron, J. (2017). The effects of marijuana liberalizations: Evidence from Monitoring the Future. *National Bureau of Economic Research*, *w23779*.

Dubé, J.-P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, *41*(3), 417–445.

Eliaz, K., & Spiegler, R. (2011). A simple model of search engine pricing. *The Economic Journal*, *121*(556), F329–F339.

Ershov, D., Laliberté, J.-W., & Orr, S. (2018). Mergers in a model with complementarity. *Scott, Mergers in a Model with Complementarity (December 3, 2018)*.

Fosgerau, M., Monardo, J., & De Palma, A. (2020). The inverse product differentiation logit model. *Available at SSRN 3141041*.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, *97*(3), 713–744.

Gilleskie, D., & Strumpf, K. (2005). The behavioral dynamics of youth smoking. *Journal of Human Resources*, *40*(4), 822–866.

Goel, R. K., & Morey, M. J. (1995). The interdependence of cigarette and liquor demand. *Southern Economic Journal*, 451–459.

Gordon, B. R., & Sun, B. (2015). A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science*, *34*(3), 452–470.

Grigolon, L., Plúas-López, A., Remmy, K., & Sovinsky, M. (2022). *High taxes: Legal marijuana and black market response* (Tech. Rep.). University of Mannheim.

Grossman, M., Chaloupka, F., & Sirtalan, I. (1998). An empirical analysis of alcohol addiction: results from the monitoring the future panels. *Economic Inquiry*, *36*(1), 39–48.

Haffajee, R., & Mauri, A. (2021). Cannabis liberalization in the us: the policy landscape. *Health Affairs Health Policy Brief*, *1*.

Hai, R., & Heckman, J. (2022). *The causal effects of youth cigarette addiction and education* (Tech. Rep.). National Bureau of Economic Research.

Hansen, B., Miller, K., & Weber, C. (2017). *The taxation of recreational marijuana: Evidence from washington state* (No. w23632). National Bureau of Economic Research Cambridge, MA, USA:.

Hansen, B., Miller, K., & Weber, C. (2020). Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. *Journal of Public Economics*, *187*, 104159.

Harding, M., Leibtag, E., & Lovenheim, M. F. (2012, May). The heterogeneous geographic and socioeconomic incidence of cigarette taxes: Evidence from nielsen homescan data.

American Economic Journal: Economic Policy, *4*(4), 169-98.

Heien, D., & Durham, C. (1991). A test of the habit formation hypothesis using household data. *The Review of Economics and Statistics*, 189–199.

Hinnosaar, M., Liu, E., & Loaeza-Albino, E. (2023). Externalities of marijuana legalization: Marijuana use in non-legalizing states. *Available at SSRN*.

Hollenbeck, B., & Giroldo, R. Z. (2022). Winning big: Scale and success in retail entrepreneurship. *Marketing Science*, *41*(2), 271–293.

Hollenbeck, B., & Uetake, K. (2021). Taxation and market power in the legal marijuana industry. *The RAND Journal of Economics*, *52*(3), 559–595.

Iaria, A., & Wang, A. (2021). An empirical model of quantity discounts with large choice sets. *CEPR Discussion Paper No. DP16666*.

Imai, S., Jain, N., & Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, *77*(6), 1865–1899.

Jacobi, L., & Sovinsky, M. (2016). Marijuana on main street? Estimating demand in markets with limited access. *American Economic Review*, *106*(8), 2009–45.

Jiang, R., Manchanda, P., & Rossi, P. E. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, *149*(2), 136–148.

Kenkel, D. (2005, May). Are alcohol tax hikes fully passed through to prices? evidence from alaska. *American Economic Review*, *95*(2), 273-277.

Kilmer, B., Everingham, S., Caulkins, J., Midgette, G., Pacula, R. L., Reuter, P., … others (2014). What america's users spend on illegal drugs: 2000–2010.

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing science*, *29*(6), 1001–1023.

Kissi, D. (2021). *'When recreational marijuana is legalized, cigarette sales rise', Chicago Booth Review.* Retrieved September 2023, from `https://www.chicagobooth.edu/review/when-recreational-marijuana-legalized-cigarette-sales-rise`.

Kosuke, I., & Van Dyk, D. (2005). A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of econometrics*, *124*(2), 311–334.

Labeaga, J. M. (1999). A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco. *Journal of econometrics*, *93*(1), 49–72.

Liu, H., Chintagunta, P., & Zhu, T. (2010). Complementarities and the demand for home broadband internet services. *Marketing Science*, *29*(4), 701–720.

Loaiza-Maya, R., & Nibbering, D. (2022). Scalable bayesian estimation in the multinomial probit model. *Journal of Business & Economic Statistics*, *40*(4), 1678–1690.

Ma, Z. (2017). The effects of increases in cigarette prices on cigarette consumption among smokers after the master settlement agreement. *Review of Economics of the Household*, *15*, 1177–1190.

Macha, V., Abouk, R., & Drake, C. (2022, 11). Association of Recreational Cannabis Legalization With Alcohol Use Among Adults in the US, 2010 to 2019. *JAMA Health Forum*, *3*(11), e224069-e224069.

McCulloch, R. E., Polson, N. G., & Rossi, P. E. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*(1),

173–193.
Melillo, G. (2022). *'As states legalized marijuana, alcohol use increased: study', Changing America.* Retrieved September 2023, from `https://thehill.com/changing-america/well-being/prevention-cures/3742213-as-states-legalized-marijuana-alcohol-use-increased-study`.
Miller, K., & Seo, B. (2021). The effect of cannabis legalization on substance demand and tax revenues. *National Tax Journal*, *74*(1), 107–145.
Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review*, *100*(4), 674.
Nelson, J., & Moran, J. (2020). Effects of alcohol taxation on prices: A systematic review and meta-analysis of pass-through rates. *The B.E. Journal of Economic Analysis & Policy*, *20*(1), 20190134. Retrieved 2023-09-27, from `https://doi.org/10.1515/bejeap-2019-0134` doi: doi:10.1515/bejeap-2019-0134
Pacula, R., Grossman, M., Chaloupka, F. J., O'Malley, P. M., Johnston, L. D., & Farrelly, M. C. (2001). Marijuana and youth. In *Risky behavior among youths: An economic analysis* (pp. 271–326). University of Chicago Press.
Pierani, P., & Tiezzi, S. (2009). Addiction and interaction between alcohol and tobacco consumption. *Empirical Economics*, *37*, 1–23.
Ritter, A., & Sotade, O. (2017). Explaining the declining rates of past year cannabis use in australia: A first pass. *Drug and Alcohol Review*, *36*(5), 602–608.
Sen, A., & Wirjanto, T. (2010). Estimating the impacts of cigarette taxes on youth smoking participation, initiation, and persistence: empirical evidence from canada. *Health economics*, *19*(11), 1264–1280.
Sovinsky Goeree, M. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, *76*(5), 1017–1074.
StatCan. (2020). *Canadian community health survey - annual component (cchs)* (Tech. Rep.). Government of Canada, Statistics Canada.
Tax Foundation. (2023). Retrieved September 2023, from `https://taxfoundation.org`.
Thomas, D. (2019). License quotas and the inefficient regulation of sin goods: Evidence from the washington recreational marijuana market. *Available at SSRN 3312960*.
Thomassen, Ø., Smith, H., Seiler, S., & Schiraldi, P. (2017). Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, *107*(8), 2308–51.
Waters, T. M., & Sloan, F. A. (1995). Why do people drink? tests of the rational addiction model. *Applied Economics*, *27*(8), 727–736.
Yu, X. (2022). Bundle choice with limited consideration: An application to yogurt.
Zhao, X., & Harris, M. N. (2004). Demand for marijuana, alcohol and tobacco: participation, levels of consumption and cross-equation correlations. *Economic Record*, *80*(251), 394–410.
Zuckermann, A., Battista, K., Bélanger, R., Haddad, S., Butler, A., & Costello, M. (2021). Trends in youth cannabis use across cannabis legalization: data from the compass prospective cohort study. *Prev Med Rep*, *22*(101).

Appendix A Data Details

A.1 Data on Marijuana Prices and Quality

For the yearly average cannabis prices per gram and the cannabis quality ratio, we use observations from two sources, the High Times and PriceOfWeed.com. Observations include amounts in eighths (3.5 grams), quarters (7 grams), half ounces (14 grams), and ounces (28 grams).

We do not observe marijuana prices and quality ratios for 20 state/year combinations. In 2004, Alaska, Idaho, Oregon, Rhode Island, Utah, Vermont. In 2007, Arkansas, North Dakota, West Virginia, Wisconsin, and Wyoming. In 2008, Alaska, Delaware, Hawaii, Kentucky, Mississippi, North Dakota. In 2009, North Dakota, West Virginia, Wyoming. Alaska and Hawaii are subsequently dropped from the sample to be consistent with the MTF sample. When this information is missing for the first year of the sample (6 observations), we assign to the state the price or quality ratio of the subsequent year. For the rest of the missings (14 observations), we employ linear interpolation. The combined dataset consists of around 139,000 transactions, from 2004 to 2013. Finally, we determine the price and quality ratio for each region/year combination, and for the price in particular we compute the weighted (by population) average over the states.

A.2 Questions on Substance Use

Substance Use Variables (Main Specification)

Our substance use variables are based on the following questions from the MTF survey. For marijuana, we use answers to the question: *"On how many occasions (if any) have you used marijuana (grass, pot) or hashish (hash, hash oil) in the past 30 days?"* Similarly, for alcohol, we use answers to the question: *"On how many occasions have you had alcoholic beverages to drink more than just a few sips during the last 30 days?"* For cigarettes, we use answers to the question: *"How frequently have you smoked cigarettes during the past 30 days?"* Our substance use variables for marijuana, alcohol and smoking are equal to one if, respectively, an individual used marijuana, drank alcohol, or smoked cigarettes in the previous month, and zero otherwise.

Simultaneous Use Variable

In the MTF survey, for a subsample of individuals, we observe whether they simultaneously use marijuana and alcohol. The survey asks the following question: *"How many of the times when you used marijuana or hashish during the last year did you use it along with alcohol– that is, so that their effects overlapped?"* The answers are *Not at all, a few of the times, some of the times, most of the times, every time.* The simultaneous use variable equals one if the respondent answers "most of the time" or "every time" to the question, and zero otherwise.

Substance Use Variables (Panel Dataset)

The substance use variables for our panel dataset are based on the following questions from the PSID survey. For marijuana, we use answers to the question: *"On how many occasions (if any) have you used marijuana in the past year?"* For alcohol, we use answers to the question: *"In the last year, on average, how often did you have any alcohol to drink?"* If an individual reports consuming marijuana on one or more occasions or drinking in the previous year, we define the individual as a current user. For cigarettes, we use answers to the question: *"Do you smoke cigarettes?"* If an individual responds positively to this question, we define the individual as a current smoker.

Appendix B Details on Estimation

B.1 MCMC Inference for MNP Substance Bundle Model

To outline the MCMC estimation strategy for the inference of the model parameters via the posterior distribution in (9), we define the latent utility vectors under both full and limited access as

$$\tilde{u}_{im} = X_{im}\Theta + I_{im}^{\nu}\nu_i + \varepsilon_{im},\ \varepsilon_{im} \sim \mathcal{N}(0, \mathbf{I}),$$

where $\Theta = (\theta_1', \theta_2', \theta_3', \gamma_4, \gamma_5, \gamma_6)'$ and the covariate matrices are defined as

$$X_{im} = \begin{pmatrix} Z_{i1m}' & 0 & 0 & 0 & 0 & 0 \\ 0 & Z_{i2m}' & 0 & 0 & 0 & 0 \\ 0 & 0 & z_{i3m}' & 0 & 0 & 0 \\ Z_{i1mt}' & Z_{i2m}' & 0 & 1 & 0 & 0 \\ Z_{i1m}' & 0 & Z_{i3m}' & 0 & 1 & 0 \\ 0 & Z_{i2m}' & Z_{i3m}' & 0 & 0 & 1 \\ Z_{i1m}' & z_{i2m}' & Z_{i3m}' & 1 & 1 & 1 \end{pmatrix}, I_{im}^{\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \text{ if } a_{im} = 1,$$

or

$$X_{im} = \begin{pmatrix} 0 & Z_{i2m}' & 0 & 0 & 0 & 0 \\ 0 & 0 & Z_{i3m}' & 0 & 0 & 0 \\ 0 & Z_{i2mt}' & Z_{i3m}' & 0 & 0 & 1 \end{pmatrix}, I_{im}^{\nu} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \text{ if } a_{im} = 0.$$

We simulate from the posterior distribution with a 4-step Gibbs-sampler that updates the latent utilities and model parameters from their corresponding full conditional posterior distribution.

1. For $i = 1, \ldots, N$, $t = 1, \ldots, T$, $r \in \mathcal{R}$ if $a_{im} = 1$ or $r \in \mathcal{R}^L$ if $a_{im} = 0$, draw

$$\tilde{u}_{irm} | \tilde{u}_{i,-r,m}, \nu_i, \Theta \quad \sim \quad \begin{cases} \mathcal{N}(\mu_{irm}, 1)\mathbf{1}(\tilde{u}_{irm} > \max(\tilde{u}_{i,-r,m}, 0)), \text{ if } y_{im} = r, \\ \\ \mathcal{N}(\mu_{irm}, 1)\mathbf{1}(\tilde{u}_{irm} < \max(\tilde{u}_{i,-r,m}, 0)), \text{ if } y_{im} \neq r, \end{cases}$$

   where $\mu_{irm} = \sum_{j \in r} Z_{ijm}\beta_j + \sum_{j \in r} \nu_{ij} + \gamma_r$.

2. Draw $\Theta | \tilde{\boldsymbol{u}}, \Sigma \sim \mathcal{N}(\bar{m}_\Theta, \bar{V}_\Theta)$ where $\bar{V}_\Theta = (\sum_i X_{im}\Omega_{im}^{-1}X_{imt}' + V_\Theta^{-1})^{-1}$ and $\bar{m}_\Theta = \bar{V}_\Theta(\sum_i \sum_t X_{im}\Omega_{imt}^{-1}\tilde{u}_{im} + V_\Theta^{-1}m_\Theta)$.

3. For $i = 1, \ldots, N$, draw $\nu_i | \tilde{\boldsymbol{u}}, \Theta, \Sigma \sim \mathcal{N}(\bar{m}_\nu, \bar{V}_\nu)$ where $\hat{u}_{im} = \tilde{u}_{im} - X_{im}\Theta$, $\bar{V}_\nu = (\sum_t I_{im}^{\nu}{}' I_{im}^{\nu} + \Sigma^{-1})^{-1}$, $\bar{m}_\nu = \bar{V}_\nu \sum_t I_{im}^{\nu}{}' \hat{u}_{im}$.

4. Draw $\Sigma | \nu \sim \mathcal{IW}(\bar{\nu}_\Sigma, \bar{\Psi}_\Sigma)$ where $\bar{\nu}_\Sigma = \nu_\Sigma + NT$ and $\bar{\Psi}_\Sigma = \Psi_\Sigma + \sum_i \nu_i \nu_i'$.

The first step simulates latent utility from a truncated normal distribution, where the truncation is determined by whether the bundle r is chosen and the highest utility among the remaining bundles $-r$ including the outside option $r = 0$. Intuitively, if a bundle r is chosen, u_{irmt} would be higher than all other utilities $u_{i,-r,mt}$ including the no-use $u_{i0mt} = 0$.

Otherwise, u_{irmt} would be lower than the utility of the chosen bundle. The second and third steps are standard normal updates. Note that we integrate out ν_i when updating Θ in the second step to improve mixing property as discussed. The last step is a standard inverted Wishart update.

To address difficulties in the estimation of the covariance of random effects, and more generally the covariance matrix in (bundle) choice models in real data settings, we scale the covariance matrix by its first element similar in spirit to the approach in McCulloch et al. (2000) for standard MNP settings. In practical terms we fix the scale-problem by scaling the Σ draws by the standard deviation of the 1st element σ_{11}. To implement this restriction, we modify the 4th step and at each iteration scale the unrestricted draw of Σ by its first element σ_{11}. We note that this is a data issue. In our simulation study, all elements of Σ can be estimated and the MCMC chain shows good convergence.

B.2 Predictive Analysis

To estimate the probabilities of substance use (10) for the counterfactual analysis and participation price elasticities we extend the Gibbs algorithm with the following prediction step after the burn-in period. We first predict the latent utility vector $\tilde{\boldsymbol{u}}^p_{im}$ based on the market and price/tax characteristics under the policy scenario (including the sample setting) and the current posterior draws of model parameters and predict the corresponding to bundle choice y^p_{im} for each individual given the relevant individual access.

> Draw $\tilde{u}^p_{im} = \{\tilde{u}^p_{irm}\}\forall i$ from $N(X^p_{im}\Theta + I^{\nu}_{im}\nu_i, \mathbf{I}_7)$ and predict bundle use $y^p_{im} = r$ as $\arg\max_{r\in(\mathcal{R}\cup 0)}\ \tilde{u}^p_{irm}\}$ if $a^p_{im} = 1$ or $\arg\max_{r\in(\mathcal{R}^L\cup 0)}\{u^p_{irm}\}$ if $a^p_{im} = 0$.

We then aggregate the predicted individual choices to compute the probability of bundle use S^p_r for $r = 0, 1, \ldots, 7$ and substance use S^p_j for $j = 1, 2, 3$ by computing

$$S^p_{rm} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(y^p_{im} = r), \text{ and } S^p_{jm} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(j \in y^p_{im})$$

using the current values of the predicted individual bundle choices. For the prediction with persistence, we apply the idea of the group-based past use in the dynamic counterfactual policy analysis and generate past use behavior based on predicted individual use within a group from the 2nd period onward. Statistics, such as the mean probability of substance use are computed from the set of draws wich simulate of the predictive use distribution.

For the distribution of the bundle and substance participation elasticities, we combine the prediction step under the status quo setting (sample values) at with two-sided numerical differentiation step. Using the set of observed prices or taxes $\{p_{ilm}\}$, we predict the bundle and substance probabilities at three points, (1) $p_{ilm}^{\text{backward}} = p_{ilm} \times (1 - 0.05)$, (2) p_{ilm}, and (3) $p_{ilm}^{\text{forward}} = p_{ilm} \times (1 + 0.05)$ and compute for example the substance elasticity as

$$\mathcal{E}_{jkm}^{p} = \frac{\partial S_j^p}{S_j^p}\frac{p_l}{\partial p_l} = \frac{S_{jm}^p(\{p_{ilm}^{\text{forward}}\}_{i=1}^N) - S_{jm}^p(\{p_{ilm}^{\text{backward}}\}_{i=1}^N)}{S_{jm}^p(\{p_{ilm}\}_{i=1}^N)}\frac{1}{10\%}$$

where we obtain the own-price elasticity if $j = l$. The above expression is evaluated at each iteration of the Gibbs sampler. Bundle participation elasticities are computed correspondingly by evaluating the predictive bundle use probabilities at each iteration.

Appendix C Estimates of Model with Persistence

Table A1: Multi-substance Use with Choice Set Restrictions and Persistence

	Limited Access (Limited Choice Set)		
	Marijuana	Cigarettes	Alcohol
Lagged Use	0.648***	0.355**	0.116
	(0.173)	(0.153)	(0.082)
Product and Market Characteristics			
Price	-0.014*	-0.088***	-0.005***
	(0.007)	(0.008)	(0.001)
High Quality Proportion	0.152		
	(0.116)		
Police Enforcement	-0.279**		
	(0.114)		
Demographics			
White	-0.269***	0.292***	0.166***
	(0.022)	(0.020)	(0.021)
Male	0.207***	0.010	-0.017*
	(0.016)	(0.010)	(0.010)
Aged 18 and over	-0.057***	0.099***	0.017
	(0.012)	(0.010)	(0.011)
Max Parental Edu: HS	-0.072***	-0.116***	0.082***
	(0.017)	(0.013)	(0.013)
Max Parental Edu: College	-0.056***	-0.236***	0.112***
	(0.016)	(0.012)	(0.012)
Bundle Effects			
With Cigarettes	0.954***		
	(0.011)		
With Alcohol	1.983***		
	(0.016)		
Cigarettes and Alcohol	0.908***		
	(0.011)		
Unobserved Heterogeneity			
Marijuana	1.000[a]	0.024***	-0.675***
Cigarettes	0.024***	0.002***	-0.016***
Alcohol	-0.675***	-0.016***	0.458***

Notes: Mean coefficients, standard deviations in parenthesis. Year fixed effects are included. Region fixed effects are included for bundles with marijuana. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The significance levels are based on the posterior credibility intervals. [a] Normalized to 1.

9 Online Appendix

9.1 Empirical Distribution Lagged Use

Table 2: Empirical Distribution Lagged Use from PSID

Demographic Groups			**Average Lagged Use**		
Over 17	Male	White	Marijuana	Cigarettes	Alcohol
No	No	No	0.18	0.08	0.23
Yes	No	No	0.17	0.10	0.35
No	No	Yes	0.33	0.21	0.58
Yes	No	Yes	0.27	0.17	0.57
No	Yes	No	0.23	0.04	0.38
Yes	Yes	No	0.29	0.13	0.45
No	Yes	Yes	0.30	0.25	0.48
Yes	Yes	Yes	0.37	0.18	0.64