Aggregative Games, Excise Taxes, and Illicit Cannabis: How Legal Regulation Fuels Illegal Markets

Braden Wagner[†]

November 6, 2026, latest version available here

Abstract

Despite the widespread legalization of cannabis, the illegal cannabis market remains prevalent throughout the United States. This paper investigates the impact of excise taxes and legal market structure on the size of the illegal cannabis market and the entry of illegal sellers, from both a theoretical and empirical perspective. I derive a unique pricing and entry equilibrium for an aggregative game in which legal firms act as dominant oligopolists, while unlicensed illegal firms act as a monopolistically competitive fringe. Next, I estimate a structural model of supply, demand, and entry on 2014–2016 Washington State data to measure the effect of excise taxes and legal market structure on the demand for illegal cannabis and to estimate the unobserved entry of illegal operations. The structure of the model provides two methods for estimating unobservable illegal entry, using a zero-profit condition. Counterfactuals show that raising the excise tax increases consumer substitution to the illegal market and encourages more illegal firms to enter. Although the excise tax aims to mitigate the externality and health costs of cannabis use, higher rates drive consumers to the unregulated illegal market, increasing those costs and lowering total welfare. Reduced-form estimates from a newly collected state panel of legal and illegal cannabis data support the structural results.

Keywords: Aggregative Games, Demand Estimation, Structural Empirical Model, Taxation and Regulation, Illicit Markets

JEL Codes: D43, K42, L00, L13, L51

I am grateful to my advisors, Federico Ciliberto, Simon Anderson, and Alexander MacKay, for their guidance and support throughout this project. I also thank the faculty, seminar participants, and graduate students who provided feedback at the University of Virginia IO workshop, weekly IO meetings, and the University of Virginia Economic Research Colloquium. Finally, I thank New Frontier Data for providing cannabis market data used in this project. All errors are my own.

[†]University of Virginia; dmj3yd@virginia.edu

1 Introduction

For centuries, underground markets for drugs and other substances have developed as a way to evade government restrictions. From the opium trade of the eighteenth and nineteenth centuries to bootlegging during Prohibition, the crack epidemic of the 1980s and today's fentanyl crisis, these illicit economies have persisted. Americans spent nearly $150 billion on illegal drugs in 2016 (Kilmer, 2019), and one report puts 2022 US cannabis sales at $106 billion, $76 billion of it illicit (New Frontier Data, 2023). Demand for illicit substances raises health concerns, and illegal drug operations also harm society and the environment. Cannabis, one of the most widely used intoxicating substances in the world, has become normalized: about 22% of Americans aged 12 and older reported past-year use in 2022, and millions use it every day (Substance Abuse and Mental Health Services Administration, 2023). It is also a clear case of an illegal market that persists next to a legal one, since illegal sellers still supply much of the market in some states that have legalized cannabis, including California and Washington (Kaste, 2024; Kilmer et al., 2019).

Illegal cannabis operations are not confined to states with no legal product. Although legalization was touted as a sure way to generate new businesses and tax revenue, it has also allowed many illegal and unlicensed operations to "hide in plain sight" within legal states. Demand for illicit cannabis is driven largely by price differences: legal producers and retailers face excise taxes and costly regulations, which in many cases raise the cost of legal cannabis well above that of illegal products. Furthermore, restrictive licensing caps and the state-fragmented legal market structure may drive more illegal entry and substitution toward unregulated products. Excise taxes are levied to curb cannabis use and raise revenue, but they also raise the cost of legal cannabis, so legal firms cannot price low enough to keep illegal sellers out. When untaxed sellers can enter freely, many consumers switch from legal stores to unregulated sellers, which in turn makes illegal entry more profitable. Cannabis use already has consequences for consumer health and the environment, and illegal, unregulated supply makes them worse and may bring other criminal activity, while the state forgoes the tax revenue on those sales. A high excise tax therefore affects not only consumer demand but also the number of illegal sellers. Measuring this response is difficult: illegal sales are observed only as aggregate estimates, and the number of illegal sellers is never observed.

This paper studies the effect of excise taxation and licensing caps within the legal cannabis market on the size of the illegal cannabis market and the entry of illegal sellers. It asks three questions. How do excise taxes and regulation affect consumer substitution between legal and illegal cannabis? How do legal market structure and taxation affect the entry of illicit operations? How does the illegal market change the welfare and revenue effects of the excise tax? To answer them, I develop an entry and pricing aggregative game in which legal firms act as dominant oligopolists and unlicensed illegal firms act as a monopolistically competitive fringe. I then estimate a structural model, using

the intuition of the aggregative-games approach to inform the empirical model with 2014–2016 data from Washington State, where recreational sales began on July 8, 2014 with a 25% retail excise tax that rose to 37% on July 1, 2015. This approach allows me to recover the unobserved number of illegal sellers and simulate counterfactual changes in the retail excise rate, with the illegal sector responding strategically to legal prices. Finally, I support the structural results with reduced-form evidence from a newly collected panel covering all 50 states and the District of Columbia, which gives a broader view of how legalization affects the illegal market and cannabis use.

The game builds on earlier aggregative-games approaches and on Forchheimer's dominant-firm model (Acemoglu and Jensen, 2013; Anderson et al., 2020; Nocke and Schutz, 2018; Selten, 1970; Forchheimer, 1908). Building on Anderson et al. (2020), I show that while illegal sellers are active, free entry holds each seller's profit at its entry cost, which fixes the outside share at a level that does not depend on any legal price, cost or tax. Under logit, a tax on legal firms therefore changes who sells cannabis but not how much is sold, and every sale the legal sector loses goes to an illegal seller. I prove that the pricing and entry equilibrium, with endogenous entry of both licensed and illegal firms, is unique under logit. In the appendix, I extend uniqueness to nested logit and to legal price leadership, where a priority rule selects among equilibria that share the same aggregate. Uniqueness under logit does not require an active fringe, so the equilibrium is determined even at taxes low enough to drive illegal sellers out. I also show that, for a given licensed set, a policymaker who ignores the harms of illegal supply sets the tax too high.

The empirical model follows the game and extends IO-standard demand and supply estimation by adding entry and indicator equations and recovering the latent level of entry at every candidate parameter vector. Demand is nested logit, legal retailers set oligopoly prices, illegal sellers set monopolistically competitive prices, and a zero-profit condition governs illegal entry; arrests for possession and for sales and manufacturing serve as noisy indicators of entry. Because the number of illegal sellers is never observed, I recover it in two ways: by substituting the zero-profit condition into the other equations, and by choosing the number of sellers that best reconciles several noisy readings of entry, four in the preferred specification (illegal demand, zero profit and the two arrest series).

Data on the Washington legal market come from administrative records kept by the state, obtained through FOIA requests and from Hollenbeck and Uetake (2021). Like most states, Washington uses "seed-to-sale" tracking, so these data are detailed: they include wholesale and retail prices, quantities, individual and aggregate sales, and product and firm characteristics from the first sales in July 2014 through 2016. For the illegal market, I use estimates from New Frontier Data and other consulting firms, prices from priceofweed.com and dark-web markets, reports from the Drug Enforcement Administration (DEA) and other law enforcement agencies, county arrest counts from the FBI Uniform Crime Reporting Program, and survey data on drug use from the National Survey

on Drug Use and Health (NSDUH) and the International Cannabis Policy Study (ICPS). Together these sources give illegal sales, prices and enforcement measures. For the reduced-form analysis, I assemble a separate panel of legal and illegal prices, sales, tax rates, dispensary counts, crime, and NSDUH use prevalence for the 50 states and the District of Columbia from 2011 through 2024.

My main finding is that once the illegal fringe is included, Washington is on the falling side of the Laffer curve. In my preferred specification, 2016 tax revenue peaks at a 20% excise rate, below Washington's 37%, whereas a legal-only model, which leaves illegal purchases in the outside good, finds revenue still rising at 37% (Figures 13 and 4). The legal-only model understates how elastic legal demand is. As with many intoxicating substances, total cannabis demand is inelastic with respect to the legal price, but demand for *legal* cannabis is much more elastic. Lowering the rate to the revenue peak would raise revenue, reduce the illegal share, and cut the number of illegal sellers. Given the external costs I assume, welfare is maximized at rates between 0% and 15%, depending on the measure, so Washington's 37% rate is above both its revenue-maximizing and its welfare-maximizing level. The mechanism can operate wherever licensed and unlicensed sellers coexist, but the estimates are for Washington alone.

Two-way fixed effects and event-study regressions on the state panel are consistent with legal and illegal sellers competing for the same consumers. Illegal prices are about 4% lower after a neighboring state opens legal sales and about 3% lower after the state itself does, and the neighbor effect is smaller the higher the neighbor's legal price. Past-year use rises by about 0.8 percentage points when a neighbor opens, while past-month use does not change significantly in that design. Because daily and near-daily users account for about 80% of cannabis consumed (Caulkins et al., 2020), a rise in past-year use without a significant rise in past-month use suggests that the increase comes mainly from less frequent users. The own-state use estimates are larger but come from a design whose event study fails the pre-trend test ($p \leq 0.001$), so I rely on the neighbor design. The panel cannot count illegal sellers, and own-state tax rates vary too little to identify most tax effects; the structural model provides both.

This paper contributes to the study of the legalization, regulation, and taxation of sin goods and of how these regulated markets interact with underground economies. Taxes, license restrictions, and other regulation shape prices, consumption and welfare in sin-good markets (Miravete et al., 2018; Conlon and Rao, 2023; Dubois et al., 2020; Hansen et al., 2026, 2020; Miller and Seo, 2021), and legalization changes use, crime and illegal prices (Miller et al., 2017; Dragone et al., 2019; Dills et al., 2017; Gavrilova et al., 2019; Adda et al., 2014; Meinhofer and Rubli, 2021; Carrieri et al., 2019; Hansen et al., 2020; Hao and Cowan, 2020). Models of the illicit side show that free entry of illegal sellers and high legal prices limit what legalization can achieve (Bertomeu, 2005; Auriol et al., 2023; Han et al., 2024), and studies that include an illegal alternative, for taxed spirits (Smith, 1976), subscription television (Rojas and Briceño, 2019) and cannabis (Jacobi

and Sovinsky, 2016; Sovinsky et al., 2024; Plúas-López, 2024), find that leaving it out changes policy conclusions. A tax can cut the output of a harmful good by more than enforcement against suppliers can, even when some producers go underground (Becker et al., 2006), and a lower-taxed alternative lowers the optimal rate, as cross-border purchases do for cigarettes (DeCicca et al., 2013). Many cannabis papers use Washington's administrative data, among them Mace et al. (2020), who estimate elasticities, incidence and deadweight loss under alternative conduct assumptions with no illicit sector. Thomas (2019) studies license quotas and lotteries, while I take the licensed set as given. Perrault (2023) models Washington's legal and illegal sectors, each as a single representative firm competing in quality and price, and Irvine and Light (2020) calibrate a nested demand model with legal and illegal cannabis for Canada, with no entry on either side. Hollenbeck and Uetake (2021) estimate pass-through and retailer market power and, as Miravete et al. (2018) do for alcohol, locate the tax on its Laffer curve. They implicitly put illegal purchases in the outside good and find that Washington is on the upward-sloping portion of the curve. My legal-only benchmark, which makes the same assumption, reproduces their conclusion (Figure 4); once illegal sellers enter as a fringe that responds to the tax, the conclusion reverses, and Washington would raise more revenue with a lower rate, not a higher one.

This paper also contributes to the theory of aggregative games and oligopolistic competition. Aggregative games and the inclusive best reply go back to Selten (1970), and most of the literature studies comparative statics with a fixed set of firms (Corchón, 1994; Acemoglu and Jensen, 2013; Nocke and Schutz, 2018). Anderson et al. (2020) add welfare, leadership and free entry of a monopolistically competitive fringe. My model updates Forchheimer's dominant firm and competitive fringe (Forchheimer, 1908), as Anderson and Bedre-Defolie (2024) and Etro (2023) do for a single dominant firm.[1] Here the dominant sector is a licensed oligopoly and the fringe is differentiated, untaxed and illegal. All firms set prices simultaneously in the main model, as in Shimomura and Thisse (2012) and Parenti (2018), so the fringe disciplines legal firms only through entry; an appendix extension gives legal firms Forchheimer's price leadership, with ranked, endogenous entry of several leaders. The equilibrium is unique under logit and nested logit, and under logit it is unique whether or not the fringe is active. Anderson et al. (2020) show that a free-entry fringe leaves the aggregate unchanged when a shock hits only the large firms; I apply this result to a tax on a licensed oligopoly, which implies that under logit, while illegal sellers are active, illegal primitives rather than the tax set the size of the market. Games with multiple leaders have received comparatively little attention; existing work studies them in other settings, such as investment and quantity competition (Aussel and Svensson, 2018; Hu and Fukushima, 2015; Mallozzi and Messalli, 2017; Alfaro and Lander, 2021; Ino and Matsumura, 2012). My model differs in that the entry of

[1] On the dominant firm and fringe, see also White (1981), Kahai et al. (1996) and Golombek et al. (2018); on games with several leaders, see Appendix A.1.2.

both leaders and followers is fully endogenous, and the price and entry equilibrium is still unique.

Finally, the paper contributes to the structural estimation of demand, supply and entry when part of the market is unobserved. Joint estimation of demand and supply by generalized method of moments (Hansen, 1982) is standard (Berry et al., 1995; Conlon and Gortmaker, 2020; Miller and Weinberg, 2017), and entry models infer fixed costs from observed numbers of firms (Berry and Waldfogel, 1999; Ciliberto and Tamer, 2009).[2] Closest in structure are equilibrium models in which a regulated sector competes with an unregulated one: informal firms (Ulyssea, 2018), peer hosts (Farronato and Fradkin, 2022), and the drug-market models of Galenianos et al. (2012) and Galenianos and Gavazza (2017), which study a market with no legal sector and, in the second paper, use free entry of drug dealers in counterfactuals. In my setting the number of illegal sellers is never observed, and it matters for the policy answer: at given shares, free entry sends a larger fraction of the sales that licensed stores lose to a tax to illegal sellers than a fixed number of sellers does (Corollary 2(iii) and the discussion after it). In the preferred specification, I recover this number in each county-year from four noisy readings (illegal demand, zero profit and two arrest series), as in multiple-indicator models of a latent variable (Jöreskog and Goldberger, 1975; Bartlett, 1937; Frey and Weck-Hannemann, 1984). The shadow-economy literature uses these models to size hidden activity; because they identify the latent variable only up to scale, its level rests on a calibration (Schneider and Enste, 2000; Fryer et al., 2013; Breusch, 2005). Recovering latent entry mixes the equations' errors, so the instruments must be valid for the combined errors, as in latent-variable instrumental variables (Bollen, 1996). Entry is recomputed at every candidate parameter vector, as mean utilities are in Berry et al. (1995), but the recovery reconciles noisy readings instead of inverting shares with a contraction mapping. Perrault (2023) do not model illegal entry; here the number of illegal sellers moves with the legal tax in the counterfactuals. The recovery method may also be useful in other illegal or informal markets whose sales can be estimated but whose sellers cannot be counted.

The rest of the paper is organized as follows. Section 2 describes the cannabis market and the data. Section 3 sets out the game between the licensed oligopoly and the illegal fringe and derives the effect of the tax. Section 4 presents the empirical model, the recovery of illegal entry and the tax counterfactuals. Section 5 reports the reduced-form evidence from the state panel, and Section 6 concludes with the policy implications, the limitations and directions for future research. The appendices contain the proofs, the price-leadership extension, and the estimates, counterfactuals and elasticities.

[2] See also MacKay and Miller (2025), Conlon and Mortimer (2013), Dubé et al. (2012), De Loecker and Scott (2026), Aguirregabiria et al. (2026), Ciliberto et al. (2021) and, for a pricing model with an unknown number of firms, An et al. (2017).

2 Background and Data

2.1 Industry Review

Millions of Americans use cannabis each year (Substance Abuse and Mental Health Services Administration, 2023), and though it is often viewed as a "harmless" substance, the cannabis industry has large consequences for consumer health and for the welfare of society and the environment, regardless of legal status. Federal restrictions on interstate commerce create fragmented state markets, since legal cannabis cannot cross state lines. In many legal states, the illegal market remains large, mainly because illegal cannabis is cheaper (Mann, 2025); in Washington's third year of licensed sales, licensed stores supplied an estimated 40–60% of the THC that residents obtained (Kilmer et al., 2019). In Washington, the average illegal price stayed near $8 per gram from 2014 to 2016 (Table 1), even as legal prices fell. Legalization changes both the legal status of cannabis use and the stigma attached to it. Police may then enforce the remaining cannabis laws less strictly (Kaste, 2024), which lets illegal firms operate openly. These illegal firms produce unregulated products, supplying the illegal market within and across state lines.

The structural estimation focuses on Washington, while the reduced-form estimation covers the entire US. I provide a timeline of Washington's legalization and tax changes in the appendix (Table 13). Before legalization, Washington's medical market was unlicensed and untracked, operating as a "gray market." In 2011 the governor, under federal pressure, vetoed dispensary licensing, and hundreds of unlicensed storefront access points opened in its place. The state left this medical market untouched during the first years of recreational legalization. Initiative 502 (2012) created separate producer, processor, and retailer licenses, and the Liquor Control Board capped retail licenses at 334 recreational stores, allocated by lottery. Licensed sales began on July 8, 2014. In 2016 the medical market was folded into the licensed system and the cap was raised to 556.[3] Because the medical market was untracked and unlicensed, I include it in the illegal sector, although it was technically legal. Washington's recreational market opened with a 25% tax rate at each of three transfer points, including wholesale.[4] On July 1, 2015, these were replaced by a single 37% excise on the retail selling price, itemized separately and excluded from the sales-tax base.[5] I estimate the effect of changing only the retail excise tax rate, leaving any applicable wholesale taxes and costs unchanged. Six map snapshots in the appendix (Figure 28) show how different forms of legalization[6] spread across the US between 2011 and 2024.

Cannabis demand is primarily recreational, with products differentiated by strain, flavor, and potency, as are alcoholic beverages. Products also differ in their effects, from pain relief to relaxation

[3] License caps did not bind: operating stores stayed below the cap in every sample year (2014–2016).

[4] I observe tax-inclusive wholesale costs, so the wholesale tax is already part of legal marginal cost.

[5] A 37% rate on the pre-tax price is 27% of the excise-inclusive shelf price, before sales tax (RCW 69.50.535).

[6] Recreational, medical, low-THC/CBD programs, or decriminalization.

or stimulation. Edibles, concentrates, and topicals offer other ways to consume cannabis. Consumers value regulated products for peace of mind, but many are price-sensitive and risk-tolerant, which leads them to substitute toward cheaper illegal options (Goodman et al., 2022). Many cannabis consumers are young (Substance Abuse and Mental Health Services Administration, 2023), and young consumers may be more price-sensitive and less averse to risk. These traits make substitution toward illegal cannabis sensitive to legal excise taxes. Firms compete on price, product quality, and product variety.

Tax structures vary across states: some states levy weight-based, potency-based, or combined producer and retail taxes (Davis et al., 2019). Municipal taxes, regulatory fees, license caps, and local bans further inflate legal cannabis prices and encourage the entry of illegal operations. Regulation, such as contaminant and potency testing, licensing, packaging guidelines, and regulated transportation, increases the marginal and fixed costs of legal firms. California lets local governments levy their own cannabis taxes (California Department of Tax and Fee Administration, 2025); Washington does not, though local sales taxes apply (Association of Washington Cities, 2021). States also mandate strict documentation and tracking to ensure compliance, such as the "seed-to-sale" tracking system that generates my legal data.

The illegal market poses substantial risks to consumer health, the environment, and society. Illegal firms have a cost advantage because they avoid the state excise taxes, licensing fees, and other costly regulations that legal firms face. They also often cut costs in criminal ways, such as diverting water (Bauer et al., 2015), using banned pesticides, processing products in ways that may expose them to mold, and exploiting labor (Mann, 2025). Illegal operations are often linked to violence and other crimes (Blood and Weber, 2024). Policymakers who ignore illegal substitution risk losing tax revenue and worsening health and environmental externalities.

2.2 Data

I construct two datasets from a wide variety of sources, one for the structural estimation and one for the reduced-form estimation. I first describe the structural dataset. For the legal market, I use FOIA requests and data from Hollenbeck and Uetake (2021) to collect product and firm characteristics, retail prices, wholesale costs, and quantities from Washington State's traceability system.[7] The data are at the transaction level, and I aggregate them to the retailer-year level; retail sales began on July 8, 2014, so the 2014 legal data cover about six months. The estimation sample has 39 county markets in each year, with 80, 193, and 336 legal retailers in 2014, 2015, and 2016, plus one aggregate illegal product per county-year. Prices are measured per gram and quantities

[7]Washington's traceability data for 2014–2016 come from BioTrackTHC; the state moved to Leaf Data Systems in 2018 and to the Cannabis Central Reporting System (CCRS) in December 2021. BioTrackTHC had fewer collection and implementation problems than the later systems. Most other states use Metrc.

in grams; legal prices are tax-inclusive shelf prices, which include the excise tax and the state and local sales taxes, and illegal prices are untaxed.

For the illegal market, I collect estimates from New Frontier Data, the DEA, law enforcement reports, consumer surveys, and crowd-sourced data. Statewide illegal sales come from New Frontier Data. I allocate the statewide total revenue to counties in proportion to each county's NSDUH cannabis-user population and, as a robustness check, in proportion to total county population. I construct illegal retail and wholesale prices from data I scraped from priceofweed.com (Davis et al., 2016) and from a dark-web dataset. The illegal product's two characteristics, its share of flower and its number of available brands, are not observed, so I fill them in with assumed values and legal averages. All other variables come from the sources above.

No source records how much cannabis a county could buy in a year, so I construct this potential quantity from survey data, as Nevo (2001) does for cereal markets. NSDUH reports the number of past-month and past-year cannabis users. I split past-month users and past-year-only users across use-frequency groups (daily, weekly, monthly, and less frequent), convert each group to use days per year, and multiply by grams per use day from Kerr and Ye (2022); Caulkins et al. (2020). Summing over groups gives annual potential grams, M_{st}. For each county-year, the outside share is the part of market size not bought from licensed retailers or illegal sellers, $s_{0st} = 1 - \left(\sum_j q_{jst} + q_{Fst}\right)/M_{st}$. Because use frequencies and grams per use day are uncertain, I check robustness to two larger market sizes: one in which every user is a daily user, and one in which past-month users are daily users and past-year-only users are monthly users. These are bounds on potential market size, not predictions of heavier use; the coefficients are stable across them.

County counts of cannabis possession arrests and of cannabis sales and manufacturing arrests come from the FBI's Uniform Crime Reporting (UCR) Program and National Incident-Based Reporting System (NIBRS). Arrests reflect enforcement effort as well as illegal activity, so the model treats them as noisy indicators of illegal entry. I parameterize illegal entry costs with police officers, lagged property crime, housing values, electricity prices, commercial land values, urban and urbanized population shares, and population density. Other county data provide controls and instruments. Table 1 gives summary statistics for the structural sample, with more in the appendix (Table 14).

For my reduced-form estimation, I construct a state-year panel dataset of illegal, medical, and recreational prices and sales for 2011–2024, covering all 50 states and the District of Columbia. I collect prices, sales, tax rates, and legal dispensary counts from state press releases, regulatory reports, and data dashboards. As with the structural model, the illegal data are collected from New Frontier Data, priceofweed.com, a dark-web dataset, and surveys. The outcomes are the log illegal retail price and NSDUH past-year and past-month use among adults 18 and older. Treatment is the first year with positive recreational retail sales in the state (own-state design) or in any neigh-

Table 1: Summary Statistics, Washington Cannabis Market, 2014–2016

Variable	**2014**	**2015**	**2016**
Excise Tax Rate	0.25	0.25 0.37	0.37
Legal Retailers	80	193	336
Average # of Brands Available	8.462 (3.912)	21.348 (9.469)	32.553 (13.861)
Average # of Products Available	26.565 (11.337)	79.921 (38.144)	119.778 (54.649)
Average Product Share of Flower	0.896 (0.083)	0.885 (0.051)	0.835 (0.054)
Average Units Sold per Retailer	16,575 (22,186)	124,782 (135,829)	201,853 (234,734)
Average Legal Retail Price	22.10 (6.24)	14.57 (2.31)	12.58 (1.81)
Average Illegal Retail Price	7.976 (0.026)	8.029 (0.035)	8.11 (0.056)
Average Wholesale Price	7.62 (1.34)	4.56 (0.61)	4.16 (0.49)
# of Flower Brands	10.100 (5.804)	27.684 (15.130)	36.292 (18.409)
# of Solid Edible Brands	2.973 (1.355)	8.457 (3.596)	12.201 (5.167)
# of Liquid Edible Brands	1.542 (0.567)	4.699 (2.022)	7.123 (2.903)
# of Extract Brands	3.408 (2.011)	18.267 (9.547)	30.655 (15.781)

Notes: Means, with standard deviations in parentheses. Two excise rates are listed for 2015 because on July 1, 2015, Washington replaced its 25% tax on retailer receipts with a 37% tax paid by consumers.

boring state (neighbor design). I include the Section 5 controls (medical legalization, population, dispensaries per capita, unemployment, and poverty), collected from government sources, and DEA enforcement measures (summary statistics in Tables 15 and 16 in the appendix).

3 Theory: Aggregative Games

I model the market as an aggregative game between a dominant legal oligopoly and a monopolistically competitive illegal fringe, building on the toolkit of Anderson et al. (2020) and its earlier drafts (Anderson et al., 2014, 2015). Each firm's payoff depends only on its own action and one aggregate of all actions, which keeps the interaction among legal firms tractable and fits atomistic illegal sellers, who cannot move the aggregate. In the main game, legal candidates enter in rank order, illegal sellers enter freely, and all active firms then set prices simultaneously (Section 3.1). The equilibrium is unique under logit (Proposition 1S), whether or not illegal sellers are active, and under nested logit, the demand system I estimate (Proposition 3S in Appendix A.2). Appendix A.1.2 treats legal price leadership (Proposition 1L) and Appendix A.2.2 gives examples for other demand systems, where uniqueness is checked numerically. The main result is in Section 3.2: while illegal sellers are active, illegal primitives alone fix the outside share, so under logit a tax on legal firms changes who sells cannabis but not how much is sold. Section 3.5 shows that, for a given licensed set, a policymaker who ignores the harms of illegal supply sets the tax too high.

3.1 Set-up and timing

Consumers. A consumer's utility from product j is $u_{ij} = v_j - p_j + \mu\varepsilon_{ij}$ and $u_{i0} = \mu\varepsilon_{i0}$, with ε i.i.d. type-I extreme value and $\mu > 0$. This is the standard logit model of product differentiation (Anderson et al., 1992), with μ measuring the degree of product heterogeneity. With a set P_L of active legal firms and a mass N of illegal sellers charging p_f,

$$s_\ell = \frac{e^{(v_\ell - p_\ell)/\mu}}{A}, \qquad s_f = \frac{e^{(v_F - p_f)/\mu}}{A}, \qquad s_0 = \frac{1}{A}, \qquad A = 1 + \sum_{k \in P_L} e^{(v_k - p_k)/\mu} + N\, e^{(v_F - p_f)/\mu},$$

and the legal and illegal sectors have shares $S_L = \sum_{\ell \in P_L} s_\ell$ and $S_F = N s_f$.

Firms. There are $\bar{L}$ licensed legal candidates, $\ell = 1, \ldots, \bar{L}$, each with one product, quality v_ℓ, marginal cost c_ℓ and fixed cost $K_\ell > 0$; active legal firms pay a per-unit excise tax τ (Washington's tax is ad valorem, and Appendix A.1.1 shows that the tax results carry over). Write $m_\ell \equiv v_\ell - c_\ell - \tau$. Illegal sellers are identical and atomistic, with quality v_F, marginal cost c_F and entry cost $K_F > 0$; they are untaxed, and $N \geq 0$ denotes the mass that enters. The subscript F marks primitives

common to all illegal sellers and f one seller's choices. Legal firms differ in quality net of marginal cost, as in the asymmetric logit oligopoly of Anderson and de Palma (2001), where higher-margin firms set higher markups and take larger shares. An active legal firm earns $(p_\ell - c_\ell - \tau)s_\ell - K_\ell$ and an illegal seller $(p_f - c_F)s_f - K_F$. Illegal sellers also avoid compliance costs, and their marginal cost may differ from legal costs.[8]

Timing.

1. *Entry.* Candidates decide in the order of their index whether to pay K_ℓ; the first candidate that declines ends the stage. Observing the legal set P_L, a mass N of illegal sellers enters freely, each paying K_F.[9]

2. *Prices.* All active firms set prices simultaneously.

3. *Purchasing.* Consumers choose a product to maximize utility.

I study subgame-perfect equilibria. Free entry in stage 1 means that an illegal seller's stage-2 operating profit equals K_F if $N > 0$ and is at most K_F if $N = 0$. A legal candidate that is indifferent stays out.

3.2 Pricing and free entry

Fix the legal set $P_L \equiv \{1, \ldots, L\}$ and the mass N. A legal firm accounts for the sales its price draws from every other product and from the outside good, so it sets the logit oligopoly markup. An illegal seller is atomistic and sets the monopolistic-competition markup, the same for every seller:

$$p_\ell - c_\ell - \tau = \frac{\mu}{1 - s_\ell}, \qquad p_f - c_F = \mu. \tag{1}$$

A small legal firm ($s_\ell \to 0$) prices almost like an illegal seller. With $p_f = c_F + \mu$, a seller earns μs_f, where $s_f = \widehat{r}_f\, s_0$ and $\widehat{r}_f \equiv e^{(v_F - c_F - \mu)/\mu}$. Free entry drives this to K_F, so whenever $N > 0$ the outside share is

$$\bar{s}_0 = \frac{K_F}{\mu\, \widehat{r}_f}. \tag{2}$$

The outside share depends only on illegal primitives; legal prices and the tax do not enter it. If legal firms pushed s_0 below $\bar{s}_0$, illegal sellers would exit; if they raised it, sellers would enter. Under logit demand, therefore, while illegal sellers are active, free entry fixes the size of the market at $1 - \bar{s}_0$,

[8] Illegal producers can cut costs in cultivation and distribution in ways that make the product less safe, though not necessarily less potent.

[9] A candidate cannot enter after a higher-ranked candidate has declined, so the legal set is always a prefix $P_L \equiv \{1, \ldots, L\}$.

and legal firms determine only its composition.[10] The markups (1) and the free-entry condition are $L + 2$ conditions linked through the logit denominator. The aggregative representation reduces them to one equation in one unknown.

3.3 Aggregative Games

In an aggregative game each firm's payoff depends only on its own action and on one aggregate of all actions, so firms interact only through the aggregate. Aggregative games and the inclusive best reply go back to Selten (1970); Nocke and Schutz (2018) and Acemoglu and Jensen (2013) study them with a fixed set of firms. The free-entry (long-run) analysis with a monopolistically competitive fringe is Anderson et al. (2020), whose notation I follow.

I write each price as an action, $a_j \equiv e^{(v_j - p_j)/\mu}$, so that a higher action means a lower price. In the estimation, $\ln a_j = \ln s_j - \ln s_0 = \delta_j$ and $\mu = 1/\alpha$, with α the price coefficient of Section 4; under nested logit (Appendix A.2) the two nests share one parameter, $\rho_L = \rho_F = 1 - \sigma$. With a mass N of illegal sellers, the aggregate is

$$A = 1 + \sum_{\ell=1}^{L} a_\ell + \int_0^N a_f \, df = 1 + \sum_{\ell=1}^{L} a_\ell + N a_f = \frac{1}{s_0}. \tag{3}$$

An illegal seller cannot move the aggregate ($\partial A / \partial a_f = 0$), so its reply $\widehat{r}_f$ solves

$$\widehat{r}_f : \frac{\partial \pi_f}{\partial a_f} + \frac{\partial \pi_f}{\partial A} \frac{\partial A}{\partial a_f} = \frac{\partial \pi_f}{\partial a_f} = 0 \tag{4}$$

An illegal seller's operating profit is $\widehat{\pi}_F(A) \equiv \mu \widehat{r}_f / A$, strictly decreasing in A, so free entry is a condition on the aggregate alone:

$$\widehat{\pi}_F(A) = K_F \quad \Longleftrightarrow \quad A = \bar{A} \equiv \frac{\mu \widehat{r}_f}{K_F} = \frac{1}{\bar{s}_0}. \tag{5}$$

Illegal sellers are identical and take identical actions. While they are active, the equilibrium aggregate is therefore $\bar{A}$ for a given entry cost, and each seller's action is $\widehat{r}_f$, whatever legal firms do. Legal firms affect only illegal entry: zero profit pins the aggregate, and market clearing then determines the mass of illegal sellers from the legal firms' actions.

A legal firm is oligopolistic. When it sets its price, N is already fixed, so a higher action (a lower price) raises A one for one ($\partial A / \partial a_\ell = 1$), and its inclusive best reply $\widetilde{r}_\ell(A)$ solves

$$\widetilde{r}_\ell(A) : \quad \frac{\partial \pi_\ell(a_\ell, A)}{\partial a_\ell} + \frac{\partial \pi_\ell(a_\ell, A)}{\partial A} \frac{\partial A}{\partial a_\ell} = \frac{\partial \pi_\ell(a_\ell, A)}{\partial a_\ell} + \frac{\partial \pi_\ell(a_\ell, A)}{\partial A} = 0. \tag{6}$$

[10] With market size M, $\bar{s}_0 = K_F / (\mu M \widehat{r}_f)$; I set $M = 1$ in this section.

Under logit, these replies satisfy the standard aggregative-game assumptions A1–A4 below (Lemma 1): profit falls in the aggregate, best replies are unique, and a legal firm's reply rises with the aggregate while its share falls, so the fixed point (7) has one solution.

Assumptions.

- **A1: Competitiveness.** For every active firm,

$$\frac{\partial \pi_j(a_j, A)}{\partial A} < 0.$$

- **A2: Unique Own-Action Optimum.** Profits are strictly quasi-concave in a_j, so the relevant first-order condition delivers a unique maximum.

- **A3: Strategic Complementarity and Positive Action.** Each legal firm's inclusive best reply $\widetilde{r}_\ell(A)$ is positive and strictly increasing in A.

- **A4: Normalized Slope Condition.** For each legal firm,

$$\widetilde{r}'_\ell(A) < \frac{\widetilde{r}_\ell(A)}{A}.$$

  Therefore, for every legal set, the normalized sum $\left(1 + \sum_{\ell \in P} \widetilde{r}_\ell(A)\right)/A$ falls in A, so the fixed point (7) has at most one root.

- **A5: Legal Rank-Ordered Entry and Priority.** Legal candidates are indexed in decreasing efficiency order, $1, \ldots, \bar{L}$, and only prefixes $P_L \equiv \{1, \ldots, L\}$ are feasible, so a lower-ranked firm cannot enter unless every higher-ranked firm has. For $\ell \in P_L$, let $\Gamma_\ell(P_L) \equiv \pi^*_\ell(P_L)/K_\ell$, where $\pi^*_\ell(P_L)$ is ℓ's operating profit in the unique (selected) continuation of P_L, and let $\Gamma_L \equiv \Gamma_L(P_L)$ be the ratio of the marginal firm.

  (a) **Entry priority.** Candidates decide in index order, each anticipating the continuation of every prefix; a decline ends the queue; and a candidate enters only if its operating profit strictly covers K_ℓ in the selected continuation of the final legal set.

  (b) **Deterrence priority.** Used only in the leader game (Appendix A.1.2). When leaders jointly limit price, the continuum of equilibrium price vectors is resolved lexicographically in rank order: the first-ranked leader receives its best attainable equilibrium ratio π_1/K_1, the next leader resolves the remaining ties, and so on. If a leader's best attainable ratio is reached at more than one action, the larger action (lower price) is selected.

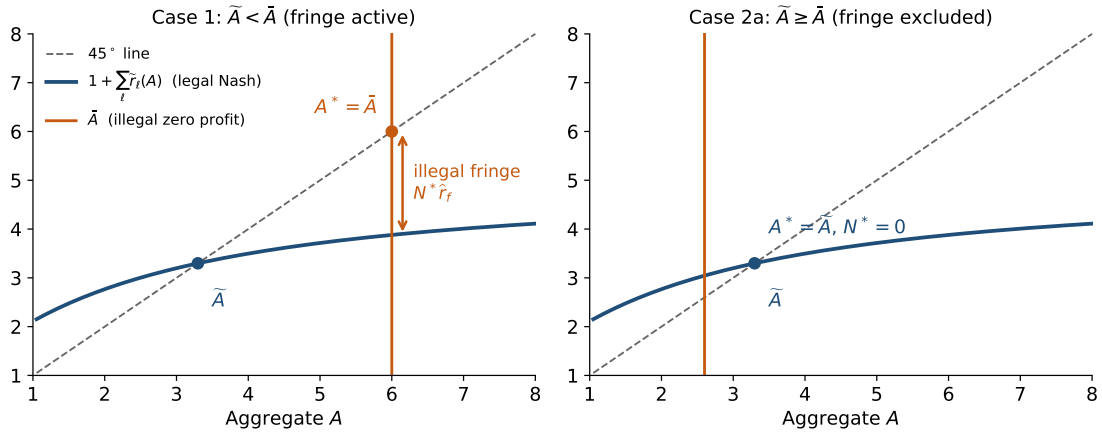


Figure 1: The equilibrium aggregate for a fixed legal set. Left: legal firms alone leave room for illegal sellers ($\widetilde{A}(P) < \bar{A}$), illegal entry fills the gap, and $A^* = \bar{A}$. Right: legal firms alone reach $\bar{A}$ ($\widetilde{A}(P) \geq \bar{A}$), so $N^* = 0$ and $A^* = \widetilde{A}(P)$. Two legal firms, $\mu = 1$.

(c) **Ordering.** $v_1 - c_1 > v_2 - c_2 > \cdots > v_{\bar{L}} - c_{\bar{L}}$ and $K_1 \leq K_2 \leq \cdots \leq K_{\bar{L}}$. Under logit this ordering implies that the ratio falls with L, $\Gamma_{L+1} < \Gamma_L$ (Lemma 3(ii); for leadership, the proof of Proposition 1L). Entry therefore stops at the first candidate whose operating profit does not cover its fixed cost.[11]

For a given legal set P and illegal mass N^*, the equilibrium aggregate solves

$$A^* = 1 + \sum_{\ell \in P} \widetilde{r}_\ell(A^*) + N^* \widehat{r}_f. \tag{7}$$

If illegal sellers are active, $A^* = \bar{A}$, each legal firm plays $\widetilde{r}_\ell(\bar{A})$, and market clearing gives

$$N^* = \frac{\bar{A} - 1 - \sum_{\ell \in P} \widetilde{r}_\ell(\bar{A})}{\widehat{r}_f}. \tag{8}$$

Let $\widetilde{A}(P)$ be the aggregate when the firms in P compete without illegal sellers, the root of $A = 1 + \sum_{\ell \in P} \widetilde{r}_\ell(A)$. If $\widetilde{A}(P) \geq \bar{A}$, legal firms alone leave no room for illegal sellers and $N^* = 0$. In every case

$$A^*(P) = \max\{\bar{A}, \widetilde{A}(P)\}, \tag{9}$$

as Figure 1 shows. Consumer surplus is $\mu \ln A$ and so depends only on the aggregate; Anderson et al. (2020) show that in a Bertrand game this property is equivalent to demand having the IIA property.

[11] For example, the set $P_{L+1;-L} \equiv \{1, \ldots, L-1, L+1\}$, which skips candidate L, is never feasible.

3.4 Equilibrium

To rule out multiple entry equilibria, candidates enter in rank order (A5(a) and the entry stage of Section 3.1) and are indexed by efficiency (A5(c)), so lower-indexed firms have "entry priority." Ranked entry follows Anderson et al. (2014), who order otherwise identical entrants by entry cost, and Anderson et al. (2015), who order firms by quality net of marginal cost.

A per-unit tax lowers every margin $m_\ell = v_\ell - c_\ell - \tau$ by the same amount, so the order does not depend on τ. For each prefix $P_L = \{1, \dots, L\}$, let $A^*(P_L)$ be the aggregate (9), and define the ratio of the marginal firm's operating profit to its fixed cost,

$$\Gamma_L \equiv \frac{\bar{\pi}_L\big(A^*(P_L)\big)}{K_L},$$

where $\bar{\pi}_\ell(A)$ is a legal firm's operating profit at its reply (Lemma 1(iv)).

Entry is resolved by rank. Potential entrants, the legal candidates and the illegal fringe, are ordered by their ratio of operating profit to fixed cost in the subgame-perfect continuation, evaluated in the final legal set. Illegal sellers earn zero profit, so the fringe's ratio is one: legal candidates with a ratio above one rank ahead of it and enter, the rest do not, and illegal sellers enter after the last legal entrant. Under logit, no tie convention is needed for illegal sellers: when the first illegal seller would exactly break even, at $\widetilde{A}(P_L) = \bar{A}$, any positive mass raises the aggregate above $\bar{A}$ and makes a loss, so $N = 0$.

The proof of Proposition 1S solves the game backward. For a fixed legal set, legal shares fall as A rises, so each illegal mass gives one aggregate. More sellers lower each seller's profit, so free entry gives one mass (Lemma 3). The entry stage is then a finite game of perfect information, and backward induction selects one legal set (Lemma 2). Assumption A5(c) determines which set it is: at any aggregate a higher-ranked firm earns more relative to its fixed cost, and each added firm weakly raises A, so Γ_L falls with L and exactly the candidates with $\Gamma_L > 1$ enter.

Proposition 1S. ***Simultaneous Equilibrium and Global Uniqueness.*** *Under logit demand (which satisfies A1–A4 by Lemma 1) and Assumptions A5(a) and A5(c), the game of Section 3.1 has a unique subgame-perfect equilibrium. Its legal set is*

$$L^* = \max\big(\{0\} \cup \{L \in \{1, \dots, \bar{L}\} : \Gamma_L > 1\}\big), \qquad P_{L^*} = \{1, \dots, L^*\},$$

its aggregate is (9)*, and its illegal mass is* (8) *if* $\widetilde{A}(P_{L^*}) < \bar{A}$ *and zero otherwise. Moreover,* $\Gamma_L > 1$ *exactly for* $L \le L^*$*, so entry stops at the first candidate with* $\Gamma_L \le 1$*. Deterrence priority A5(b) is not needed.*

Proof. See Appendix A.1. □

Under logit, the case that matters for the data is $\widetilde{A}(P_{\bar{L}}) < \bar{A}$: the licensed candidates together leave room for illegal sellers. Since $\widetilde{A}$ rises with the legal set (Lemma 3(i)), the aggregate is then $\bar{A}$ whatever subset of them enters. With illegal sellers active, each legal firm's price and profit depend only on its own margin and on $\bar{A}$, so $\Gamma_\ell = \bar{\pi}_\ell(\bar{A})/K_\ell$ whichever other candidates enter. Entry is then a dominant-strategy decision. The legal set is $\{\ell : \bar{\pi}_\ell(\bar{A}) > K_\ell\}$ whether candidates decide one at a time in any order or all at once, and this does not require A5(c); under A5(c) the set is a prefix and is also the outcome of rank-ordered entry. In the case that matters for the data, the ranking assumptions therefore do no work in selecting the equilibrium. Under nested logit, a legal firm's profit depends on the legal set even with illegal sellers active (Remark 3), but the counterfactuals hold the legal set fixed.

3.5 Welfare with Externalities

Anderson et al. (2020) show that if demand exhibits the IIA property, then consumer welfare depends only on the size of the aggregate and not its composition. In the cannabis market, the composition matters to policymakers. Besides raising revenue and curbing the externalities of legal use, they should care about the externalities of the illegal market and the tax revenue lost when consumers substitute toward illegal products. Under logit, consumer surplus still depends only on the aggregate, but total welfare also depends on its legal and illegal components, $A_L = \sum_{\ell=1}^{L} \widetilde{r}_\ell(A)$ and $A_F = N\widehat{r}_f$, so that $A = 1 + A_L + A_F$.

A policymaker sets the per-unit excise tax τ paid by legal firms, whose payoff is $\pi_\ell = (v_\ell - \mu \ln a_\ell - c_\ell - \tau)a_\ell/A$. Total welfare is

$$\mathrm{TW}(A_L, A_F, A, N, \tau) = \mathrm{CS}(A) + \mathrm{PS}(A, \tau) + \mathrm{TR}(A_L, \tau) - C_L(A_L) - C_F(A_F) - C_F^E(N),$$

where CS is consumer surplus, PS is producer surplus, TR is tax revenue, C_L and C_F are the external costs of the legal and illegal sectors, as functions of A_L and A_F, and $C_F^E(N)$ is the external cost of illegal entry. I assume that external costs are strictly increasing and convex and that $C_L(A_L) \leq C_F(A_F)$ when $A_L = A_F$. Some states may set tax rates without counting the external costs of the illegal market. I call such a policymaker *naive*: it maximizes total welfare without the costs of illegal supply and illegal entry,

$$\mathrm{NW}(A_L, A, \tau) = \mathrm{CS}(A) + \mathrm{PS}(A, \tau) + \mathrm{TR}(A_L, \tau) - C_L(A_L).$$

Proposition 2. ***Total Welfare with Externalities.*** *Although IIA demand implies that consumer welfare depends only on the size (and not the composition) of the aggregate, when firms have negative externalities, total welfare depends on the size and composition of the aggregate.*

Proof. See Appendix A.1. □

Let τ^* maximize total welfare TW and τ^n maximize naive welfare NW. A naive policymaker sets the tax too high: for a fixed nonempty legal set $\tau^n \geq \tau^*$, strictly if τ^* is interior (Appendix A.1). Lowering the tax to τ^* raises legal sales and lowers illegal entry; under logit the aggregate is unchanged (Lemma 4), and under nested logit illegal entry still falls (Corollary 2(iv)). Under logit every sale the legal sector loses goes to illegal sellers (Lemma 4(ii)); under nested logit a fraction $S_F/(S_F + \sigma s_0)$ does (Corollary 2(iii)), about 60% at the 2016 estimates in Section 4, and the rest leaves the market. Proposition 2 builds on Anderson et al. (2020): with a monopolistically competitive fringe active before and after a change that affects only the oligopolists, the aggregate is unchanged and, when consumer welfare depends only on the aggregate, total welfare changes only through the affected firms' rents. They note that this neutrality fails when the composition of the aggregate matters. Proposition 2 shows this failure for a tax on legal firms when legal and illegal supply cause different harms.

4 Empirical Model

I estimate a structural model of a dominant *legal* oligopoly and an *illegal* monopolistically competitive fringe, with endogenous prices and fringe size. The empirical model is the nested logit version of the game set up in Section 3.1, estimated on the Washington data. My main empirical contribution is to estimate an oligopoly facing a monopolistically competitive fringe whose size is never observed. I observe total illegal sales in each county-year, but not the number of illegal sellers. Legal retailers set Nash prices taking the mass of illegal sellers as given; illegal sellers charge a constant markup and enter until profit equals entry cost. I estimate demand, supply and entry jointly. As in MIMIC estimation,[12] I add indicator equations, which supply more moments that carry information on illegal entry. I recover N_{st} in two ways: by substituting the zero-profit condition into every equation that contains N_{st}, and by recovering $\ln N_{st}$ at every candidate parameter vector as the value that best reconciles the noisy readings of entry (Section 4.3). Because N_{st} responds to policy, the counterfactual taxes move illegal entry as well as illegal sales.

4.1 Demand

Consumer demand for legal and illegal cannabis is modeled as a nested logit, with three nests: legal, illegal and the outside good. Consumer i's utility from product j in nest g, bought in county

[12] Multiple indicators, multiple causes (MIMIC) models estimate a latent variable from several noisy indicator equations (Jöreskog and Goldberger, 1975).

s in year t, is

$$u_{ijst} = x'_{jst}\beta - \alpha p_{jst} + \xi_{jst} + \nu_{i\,g(j)\,st} + (1-\sigma)\varepsilon_{ijst} \tag{10}$$

The characteristics x_{jst} are the product share of flower and the number of brands available, and p_{jst} is the price of product j. For the illegal product the two characteristics are set to legal averages, and illegal prices and sales come from the sources in Section 2.2. The demand shock is ξ_{jst}, ε_{ijst} is type I extreme value, and $\nu_{i\,g(j)\,st}$ is a taste common to all products in nest $g(j)$. I let $\delta_{jst} = x'_{jst}\beta - \alpha p_{jst} + \xi_{jst}$. When $\sigma = 0$, the model reduces to multinomial logit (MNL).

I observe each legal retailer's sales but only total illegal sales in each market. Following the theory, illegal sales are supplied by a mass N_{st} of identical sellers, each with share S_{Fst}/N_{st} of the market and share $1/N_{st}$ of the illegal nest. Inverting the nested logit shares (Berry, 1994) gives one equation for each legal retailer and one for the illegal composite:

$$\ln\frac{s_{jst}}{s_{0st}} = x'_{jst}\beta - \alpha p_{jst} + \sigma \ln s_{jst|L} + f^D_{rt} + \xi_{jst}, \qquad j \in \{1,\ldots,L\}, \tag{11}$$

$$\ln\frac{S_{Fst}}{s_{0st}} = (1-\sigma)\ln N_{st} + x'_{Fst}\beta - \alpha p_{Fst} + \phi + f^D_{rt} + \xi_{Fst}, \tag{12}$$

where $s_{jst|L} = s_{jst}/S_{Lst}$ is retailer j's share of legal sales and ϕ is a constant for the illegal product. Equation (12) comes from inverting the share of one illegal seller, $\ln(s_{fst}/s_{0st}) = \delta_{Fst} + \sigma\ln(1/N_{st})$, and adding $\ln N_{st}$ to both sides. More sellers raise illegal sales, but by less when σ is large, because consumers who buy illegally substitute mostly among illegal sellers. Because only aggregate illegal sales are observed, the illegal equation has no within-nest share term. Instead, $\ln N_{st}$ enters with coefficient $1-\sigma$, which equals one under MNL; when σ is high, entry explains less of the mean utility of illegal demand. The parameter σ also enters both markups and, through the illegal markup, the zero-profit condition.

4.2 Supply

On the illegal side, sellers form a monopolistically competitive fringe of identical, single-product firms that set prices by their first-order condition and enter until profit is zero. Each seller solves

$$\max_{p_f}\ (p_f - c_{Fst})\, M_{st}\, s_{fst}(p_f, p_{-f}, N_{st}) - K_{Fst}, \tag{13}$$

where $s_{fst} = S_{Fst}/N_{st}$ at symmetric prices, c_{Fst} is the illegal marginal cost and K_{Fst} the entry cost, both common to the sellers in market st. Illegal sellers pay neither the excise nor the sales taxes.

Illegal prices are symmetric within a county-year. Because each seller is atomistic, its first-order condition gives $p_{Fst} = c_{Fst} + \frac{1-\sigma}{\alpha}$. Free entry sets profit to zero:

$$(p_{Fst} - c_{Fst}) M_{st} s_{fst}(p_f, p_{-f}, N_{st}) - K_{Fst} = 0 \tag{14}$$

The number of sellers N_{st} is the value at which this condition holds. Because I observe neither illegal entry nor entry costs, I impose two restrictions on costs: entry costs and marginal costs are the same for all illegal sellers in a market, and the entry-cost level K_0 is fixed, which pins down the level of N_{st}.

Legal retailers are single-product oligopolists. Each legal retailer solves

$$\max_{p_{jst}} \left((1 - \tau_{st}) p_{jst} - c_{jst}\right) s_{jst}(X, P, N) \tag{15}$$

Here, p_{jst} is the price of legal product j, measured as the full customer payment, and τ_{st} is the tax wedge: the share of that payment that goes to the excise and the state and local sales taxes, at combined sales-tax rate t, so the retailer keeps $1 - \tau_{st}$. Before July 2015 the 25% excise was a tax on the retailer's receipts, so $1 - \tau_{st} = 0.75/(1 + t)$; from July 2015 the 37% excise is added to the price outside the sales-tax base, so $1 - \tau_{st} = 1/(1.37 + t)$; 2015 combines the two halves using the state's reported sales and excise receipts (Appendix A.3.6). Oligopoly pricing gives $p_{jst} = \frac{c_{jst}}{1-\tau_{st}} + \frac{1-\sigma}{\alpha\left(1-\sigma s_{jst|L}-(1-\sigma)s_{jst}\right)}$. Under MNL the markup term is $1/[\alpha(1 - s_{jst})]$. Log marginal cost is

$$\ln c_{jst} = w'_{jst}\gamma + f^S_{rt} + \eta_{jst}. \tag{16}$$

4.3 Entry: Zero Profit, Arrest Indicators, and Recovering N_{st}

The number of illegal sellers N_{st} is not observed, but it appears in four equations: illegal demand, the zero-profit condition and two arrest equations. Illegal demand is Equation (12). The zero-profit condition uses the monopolistically competitive markup:

$$\frac{1 - \sigma}{\alpha} \frac{S_{Fst}}{N_{st}} M_{st} = K_{Fst}, \qquad \ln K_{Fst} = \ln K_0 + W'_{Fst}\theta + \zeta_{Fst}, \tag{17}$$

where K_0 is the entry-cost level, fixed in estimation, W_{Fst} are the entry-cost shifters and ζ_{Fst} is the entry-cost shock.

Cannabis sales and manufacturing arrests ($g = 1$, subscript s in the moment blocks and tables) and possession arrests ($g = 2$, subscript p) rise with the number of sellers but also with enforcement:

$$y_{gst} = \lambda_g \ln N_{st} + C'_{st}\psi_g + f^g_{rt} + \iota_{gst}, \qquad g = 1, 2, \tag{18}$$

where the controls C_{st}, the logs of lagged other-drug incidents and of police officers, enter only in Models 8 and 9.

Moving every term except $\ln N_{st}$ and the shock to the left of each equation turns it into a noisy reading of entry: d_{st} from illegal demand, z_{st} from zero profit and y^*_{gst} from each arrest equation. Each left-hand side depends only on data and parameters:

$$\begin{aligned} d_{st} &\equiv \ln \frac{S_{Fst}}{s_{0st}} - x'_{Fst}\beta + \alpha p_{Fst} - \phi - f^D_{rt} && = (1-\sigma)\ln N_{st} + \xi_{Fst}, \\ z_{st} &\equiv \ln \frac{1-\sigma}{\alpha} + \ln M_{st} + \ln S_{Fst} - \ln K_0 - W'_{Fst}\theta && = \ln N_{st} + \zeta_{Fst}, \\ y^*_{gst} &\equiv y_{gst} - C'_{st}\psi_g - f^g_{rt} && = \lambda_g \ln N_{st} + \iota_{gst}. \end{aligned} \tag{19}$$

Each reading is entry times a loading, plus a shock. No single reading reveals $\ln N_{st}$, but together they discipline it, as in multiple-indicator models (Jöreskog and Goldberger, 1975). I recover entry in two ways.

The first route is substitution (Models 4–6): zero profit alone gives $\ln N_{st} = z_{st} - \zeta_{Fst}$. Substituting this into illegal demand replaces $(1-\sigma)\ln N_{st} + \xi_{Fst}$ with $(1-\sigma)z_{st} + [\xi_{Fst} - (1-\sigma)\zeta_{Fst}]$, and substituting it into each arrest equation replaces $\lambda_g \ln N_{st} + \iota_{gst}$ with $\lambda_g z_{st} + [\iota_{gst} - \lambda_g \zeta_{Fst}]$. Entry drops out, but the entry-cost shock enters the bracketed errors (Appendix A.3.2).

The second route, concentration (Models 7–9), keeps zero profit as its own equation and recovers $\ln N_{st}$ as the value that best reconciles the four readings. Given the parameters and positive weights $\omega_D, \omega_Z, \omega_1, \omega_2$, it chooses $\ln N$ to minimize the weighted squared gaps between each reading and what entry predicts for it,

$$\omega_D \left[d_{st} - (1-\sigma)\ln N\right]^2 + \omega_Z \left[z_{st} - \ln N\right]^2 + \textstyle\sum_g \omega_g \left[y^*_{gst} - \lambda_g \ln N\right]^2. \tag{20}$$

The minimizer is

$$\widehat{\ln N_{st}} = \frac{\omega_D(1-\sigma)\, d_{st} + \omega_Z\, z_{st} + \omega_1\lambda_1\, y^*_{1st} + \omega_2\lambda_2\, y^*_{2st}}{\Delta}, \tag{21}$$

with $\Delta = \omega_D(1-\sigma)^2 + \omega_Z + \omega_1\lambda_1^2 + \omega_2\lambda_2^2$. Model 7 sets $\omega_1 = \omega_2 = 0$. Write $a_D = \omega_D(1-\sigma)/\Delta$, $a_Z = \omega_Z/\Delta$ and $a_g = \omega_g\lambda_g/\Delta$ for the weight on each reading. Substituting the readings (19) into (21) shows that recovered entry is true entry plus a weighted combination of the four shocks:

$$\widehat{\ln N_{st}} = \ln N_{st} + \bar{e}_{st}, \qquad \bar{e}_{st} \equiv a_D\, \xi_{Fst} + a_Z\, \zeta_{Fst} + a_1\, \iota_{1st} + a_2\, \iota_{2st}. \tag{22}$$

Putting $\widehat{\ln N}_{st}$ back into each reading gives the residuals that enter the moments:

$$\begin{aligned}
\widehat{\xi}_{Fst} &\equiv d_{st} - (1-\sigma)\,\widehat{\ln N}_{st} &&= \xi_{Fst} - (1-\sigma)\,\bar{e}_{st}, \\
\widehat{\zeta}_{Fst} &\equiv z_{st} - \widehat{\ln N}_{st} &&= \zeta_{Fst} - \bar{e}_{st}, \\
\widehat{\iota}_{gst} &\equiv y^{*}_{gst} - \lambda_g\,\widehat{\ln N}_{st} &&= \iota_{gst} - \lambda_g\,\bar{e}_{st}.
\end{aligned} \tag{23}$$

For example, with MNL demand ($\sigma = 0$), equal weights and only d and z, recovered entry is $(d_{st} + z_{st})/2$ and the illegal-demand residual is $(\xi_{Fst} - \zeta_{Fst})/2$. Entry cancels, but each residual now mixes the shocks. Substitution is the special case $a_Z = 1$ and $a_D = a_1 = a_2 = 0$, which gives the bracketed errors above. The instruments must therefore be valid for these combined errors, not only for each equation's own shock (Bollen, 1996): a demand instrument uncorrelated with ξ_{Fst} but correlated with the entry-cost shock ζ_{Fst} is invalid on either route. I assume that each equation's excluded instruments are uncorrelated, after fixed effects and controls, with every shock in its combined error. Correlation among the shocks themselves does not invalidate the moments.

Recovery chooses entry to reconcile the readings, not to minimize the GMM criterion. The estimator therefore recomputes $\widehat{\ln N}_{st}(\Theta)$, which has the closed form in Equation (21), at every candidate value of the full parameter vector Θ and minimizes

$$Q_P(\Theta) = g\big(\Theta, \widehat{\ln N}(\Theta)\big)'\, W_G\, g\big(\Theta, \widehat{\ln N}(\Theta)\big), \tag{24}$$

where $g(\Theta, \ln N)$ stacks the moments (34) with entry set to $\ln N$, and W_G is the GMM weighting matrix. Entry is never held fixed while the parameters are fitted. This resembles BLP (Berry et al., 1995), where mean utilities are recovered inside the objective; the difference is that entry is reconciled from noisy readings rather than inverted exactly from shares. Appendices A.3.3–A.3.5 give the case with cross weights, the weights of each model and the computation.

4.4 System of Equations

The full estimating system is below. Marginal costs depend on the cost shifters w_{kst} (the wholesale price, the number of available brands and the flower share) and illegal entry costs on W_{Fst} (police officers, lagged property crime, housing values, electricity prices, commercial land values per acre, the urban and urbanized shares of the population, and population density); η and ζ are cost shocks. I use NSDUH region and year fixed effects, f_{rt}, rather than county effects, to preserve county-level instruments. Without N_{st} and the zero-profit row, the first four rows are a standard

demand-and-supply system.

$$\text{Legal demand:} \quad \ln \frac{s_{jst}}{s_{0st}} = x'_{jst}\beta - \alpha p_{jst} + \sigma \ln s_{jst|L} + f^D_{rt} + \xi_{jst}, \qquad j = 1, \ldots, L, \tag{25}$$

$$\text{Illegal demand:} \quad \ln \frac{S_{Fst}}{s_{0st}} = (1-\sigma)\ln N_{st} + x'_{Fst}\beta - \alpha p_{Fst} + \phi + f^D_{rt} + \xi_{Fst}, \tag{26}$$

$$\text{Legal supply:} \quad p_{jst} = \frac{c_{jst}}{1-\tau_{st}} + \frac{1-\sigma}{\alpha\left(1-\sigma s_{jst|L} - (1-\sigma)s_{jst}\right)}, \qquad j = 1, \ldots, L, \tag{27}$$

$$\text{Illegal supply:} \quad p_{Fst} = c_{Fst} + \frac{1-\sigma}{\alpha}, \tag{28}$$

$$\text{Costs:} \quad \ln c_{kst} = w'_{kst}\gamma + \gamma_F \, \mathbb{1}\{k = F\} + f^S_{rt} + \eta_{kst} \quad (k = j, F),$$

$$\text{Entry cost:} \quad \ln K_{Fst} = \ln K_0 + W'_{Fst}\theta + \zeta_{Fst},$$

$$\text{Zero profit:} \quad \ln N_{st} = \ln \frac{1-\sigma}{\alpha} + \ln M_{st} + \ln S_{Fst} - \ln K_{Fst}, \tag{29}$$

$$\text{Arrest indicators:} \quad y_{gst} = \lambda_g \ln N_{st} + C'_{st}\psi_g + f^g_{rt} + \iota_{gst}, \qquad g = 1, 2, \tag{30}$$

$$\text{Concentrated entry:} \quad \widehat{\ln N_{st}} = \frac{\omega_D(1-\sigma)\, d_{st} + \omega_Z\, z_{st} + \omega_1\lambda_1\, y^*_{1st} + \omega_2\lambda_2\, y^*_{2st}}{\omega_D(1-\sigma)^2 + \omega_Z + \omega_1\lambda_1^2 + \omega_2\lambda_2^2}. \tag{31}$$

The readings d_{st}, z_{st} and y^*_{gst} are those of Equation (19), evaluated at the current parameters. Equation (31) is the recovery (21) with diagonal weights in Models 7 and 8; Model 9 uses a full weighting matrix (Appendix A.3.3). The zero-profit row carries no fixed effects. Figure 2 summarizes the two routes.

With K_0 fixed, I estimate nine models, grouped by how they treat entry (Table 2).

The benchmarks ignore the zero-profit condition and drop N_{st}, using only demand, or demand and supply. The substitution models add the entry equations and substitute the zero-profit condition for $\ln N_{st}$ in each equation that contains it. The concentration models use different sets of equations but do not substitute N_{st} out. Instead, they recover $\ln N_{st}$ from Equation (31) at every candidate parameter vector.

The benchmarks show what a standard demand-and-supply system, or one that leaves illegal sales in the outside good, would conclude. The substitution models show what zero profit delivers on its own. Model 4 holds the entry cost constant, Model 5 adds the entry-cost shifters, and Model 6 adds the two arrest equations. Models 8 and 9 add the arrests to the concentration route and show how much the results depend on how the readings are weighted. Model 7 attributes the common movement of its two readings to entry at each candidate parameter vector and weights each reading by the inverse of what remains. Model 8 fixes the four weights at values calibrated once. Model 9 uses a full working covariance, estimated once from a first-step fit and then held fixed, so correlated noise across readings enters the reconciliation. In every model the weights are a working choice that decides how competing readings are reconciled; they are not an estimate of the shocks' covariance

Table 2: The nine estimated models. Models 1–3 omit illegal entry, Models 4–6 substitute the zero-profit condition for N_{st}, and Models 7–9 recover $\ln N_{st}$ at every candidate parameter vector from the readings of illegal demand and zero profit (Model 7) or of these and the two arrest series (Models 8 and 9). The second column gives the rows of the system, Equations (25)–(31), that each model uses.

Model	Rows used	Illegal entry N_{st} from
Benchmarks without entry		
Demand only (Model 1)	(25), (26)	—
No entry (Model 2)	(25)–(28)	—
Legal only (Model 3)	(25), (27)	— (illegal in outside good)
Substitution route		
Substitution, constant entry cost (Model 4)	(25)–(29), K_{Fst} constant	zero profit
Substitution, entry-cost shifters (Model 5)	(25)–(29)	zero profit
Substitution with arrests (Model 6)	(25)–(30)	zero profit
Concentration route		
Concentration, two readings (demand and zero profit) (Model 7)	(25)–(29), (31)	two readings
Concentration, four readings, fixed weights (Model 8)	(25)–(31)	four readings, fixed weights
Concentration, four readings, covariance weights (Model 9)	(25)–(31)	four readings, covariance weights

Notes: Model numbers in small type are those used in the appendix tables and figure legends. Models 1–3 drop $\ln N_{st}$ from Equation (26). Every model is estimated with MNL ($\sigma = 0$) and with nested logit demand.

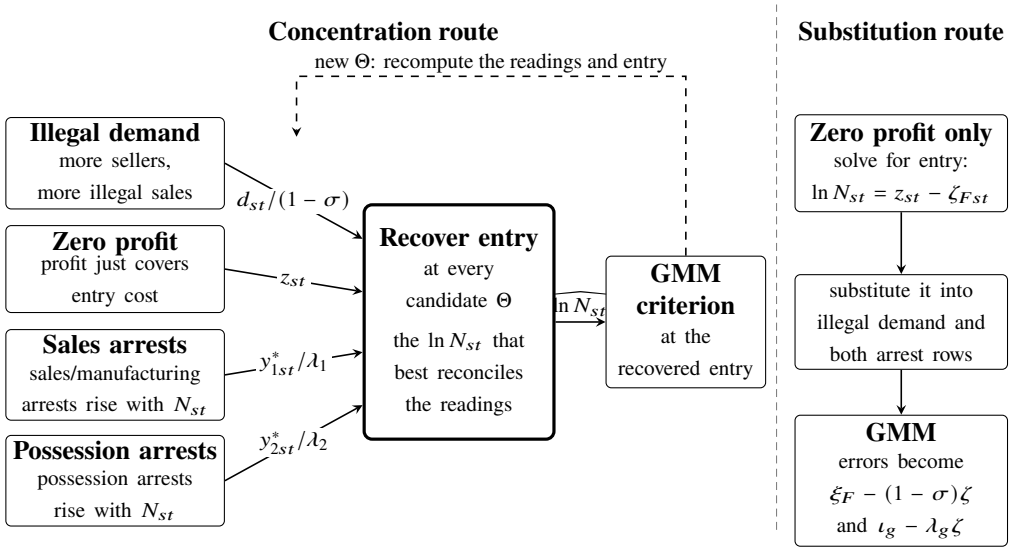


Figure 2: Recovering the number of illegal sellers. Left, Models 7–9: each equation gives a noisy reading of $\ln N_{st}$ (two readings in Model 7, four in Models 8 and 9); at every candidate parameter vector, entry is recovered as the weighted combination that best reconciles the readings, and the GMM criterion is evaluated there. Right, Models 4–6: zero profit alone is substituted for $\ln N_{st}$, so the entry-cost shock enters the illegal-demand and arrest errors.

(Appendix A.3.4).

4.5 Estimation

I estimate the parameters by IV-GMM. Let Z denote the instruments,[13] $(\xi, \eta, \iota, \zeta)$ the structural shocks, and $g(\Theta)$ the stacked moment conditions:

$$\widehat{\Theta} = \arg\min_{\Theta} \; g(\Theta)' W_G g(\Theta) \tag{32}$$

For the substitution route the moments are

$$g^{\text{sub}}(\Theta) = \begin{bmatrix} \widetilde{Z}'_{D,L}\,\xi_L \\ \widetilde{Z}'_{D,F}\,(\xi_F - (1-\sigma)\zeta) \\ \widetilde{Z}'_{S,L}\,\eta_L \\ \widetilde{Z}'_{S,F}\,\eta_F \\ \widetilde{Z}'_{I_s}\,(\iota_s - \lambda_s\zeta) \\ \widetilde{Z}'_{I_p}\,(\iota_p - \lambda_p\zeta) \end{bmatrix}. \tag{33}$$

[13]The instruments are the flower share, the wholesale price, the number of brands, the county vote share for legalization (Initiative 502), weather, the differentiation instruments of Gandhi and Houde (2019), the license cap, the entry-cost shifters W_{Fst} and market size.

There is no separate block for the zero-profit condition: its shock ζ sits inside the illegal-demand and arrest residuals, so the instruments in those blocks must be valid for the composite errors. Models 4 and 5 omit the last two blocks; Model 1 (Demand only) uses the two demand blocks and Model 2 (No entry) the first four, both with ξ_F in place of the composite error; Model 3 (Legal only) uses the two legal blocks. Throughout, $\widetilde{Z}$ denotes the instruments after the fixed effects and controls of each row are partialled out.

For the concentration route the moments are evaluated at recovered entry $\widehat{n}_{st}(\Theta) \equiv \widehat{\ln N}_{st}(\Theta)$ from Equation (21), recomputed at every candidate Θ, so the residuals below are the combined errors (23):

$$g^{\text{con}}(\Theta) = \begin{bmatrix} \widetilde{Z}'_{D,L}\,\xi_L \\ \widetilde{Z}'_{D,F}\,\xi_F(\widehat{n}) \\ \widetilde{Z}'_{S,L}\,\eta_L \\ \widetilde{Z}'_{S,F}\,\eta_F \\ \widetilde{Z}'_{E}\,\zeta(\widehat{n}) \\ \widetilde{Z}'_{I_s}\,\iota_s(\widehat{n}) \\ \widetilde{Z}'_{I_p}\,\iota_p(\widehat{n}) \end{bmatrix}, \qquad \begin{aligned} \xi_{Fst}(\widehat{n}) &= d_{st} - (1-\sigma)\,\widehat{n}_{st}, \\ \zeta_{Fst}(\widehat{n}) &= z_{st} - \widehat{n}_{st}, \\ \iota_{gst}(\widehat{n}) &= y^*_{gst} - \lambda_g\,\widehat{n}_{st}. \end{aligned} \tag{34}$$

Tables 5–7 (nested logit) and 8–10 (MNL) in the appendix report the estimates for every specification. The models largely agree on the most important parameters, the price coefficient and the wholesale-cost coefficient, and the estimates are robust to reasonable changes in the instrument set. The Legal only model gives the lowest price sensitivity under both demand systems. Under MNL, the concentration models recover fewer illegal sellers than the substitution models; under nested logit, the two routes give numbers in a similar range. The entry estimates are the least precise part of the model, and more data or less noisy indicators would sharpen them. Section 4.6 uses the estimates to simulate how tax changes affect the demand for illicit cannabis, the entry of illegal sellers, revenue and welfare.

The nested and MNL estimates differ in four ways. First, the illegal constant ϕ is negative in every nested model, including Demand only and No entry, whereas under MNL it is negative only in the models with entry. Second, because σ is high, legal and illegal markups are both smaller under nesting, so prices respond more to the tax in the counterfactuals, and substitution within a nest is much stronger than substitution across nests. Third, the smaller illegal markup lowers estimated illegal entry. Fourth, the price coefficient α is larger on average under nesting.

The estimate of σ is high because of how the nests are drawn. Each county-year has a single legal nest that contains every store, so a store's within-nest share $s_{j|L}$ is small, and smallest in the large counties with the most stores. The within-nest share therefore moves closely with the number of stores and with county population, and σ partly picks up market size. The high σ gives extreme store-level elasticities, while the legal-sector, illegal-sector and all-cannabis elasticities are

in a plausible range (Table 3). Through the pricing condition, a store's margin on the revenue it keeps is $-100/\varepsilon_{jj}$, so the implied median margin is between 1% and 5% across the nested models (Table 3). Retail margins in Washington are far higher: about 50% over wholesale (Hollenbeck and Uetake, 2021), and BOTEC's benchmarks for a typical retailer imply a gross margin of about 60% and a net margin of about 13% after compliance costs (Zamarra, 2013). Operating costs such as labor and testing explain part of the gap, and the high σ explains the rest. Stores compete more locally than across a whole county, and city or zip-code markets would give better-specified nests, but the illegal data are county counts or statewide totals allocated to counties, so they cannot be split further.

I check this in three ways (Tables 18–20). First, I split the legal nest into city nests within each county-year and keep one illegal nest, so that a store competes mainly with stores in its own city. This lowers σ to between 0.2 and 0.5 and makes the store elasticities far less extreme, but it probably lowers σ too far, because the illegal sector stays a single county-wide nest. Second, I add log population to demand, which controls for the larger number of stores in large counties, though imperfectly. Third, I restrict the sample to 2015–2016. Each change lowers σ in at least some models, and the city nests lower it the most, while the sector and all-cannabis elasticities move much less than the store elasticities. With city-year or city-month markets, I would expect σ between 0.4 and 0.7. Across all of these specifications σ stays well above zero, and nesting keeps the illegal constant ϕ negative in every model, so I keep the nested model, not the MNL, as the main specification.

Two other choices matter little. The first is market size. With either alternative bound on M_{st} from Section 2.2, every user a daily user or past-month users daily and past-year-only users monthly, the price, cost and entry-cost coefficients change little and the tax results keep the same shape. The second is the entry-cost level K_0, which I fix rather than estimate. Raising K_0 by a factor e^c lowers every $\ln N_{st}$ by c; the illegal-demand constant and the arrest constants absorb the shift, and every other estimate is unchanged (Appendix A.3.2). The level of illegal entry is therefore conditional on K_0, but the ratios of N across counties and years, the percentage response of N to a tax, and the price, quantity and revenue counterfactuals are not.

4.6 Elasticities, Counterfactuals, and Welfare

I vary the retail excise rate from 0% to 60%, holding the licensed set, the state sales tax and local taxes fixed, for every model except Demand only (Model 1), which has no pricing condition. At each rate I re-solve prices, quantities and illegal entry. Illegal sellers break even at the new prices; this is the equilibrium of Proposition 3S and Corollary 2. Under MNL, free entry pins the outside share, as Lemma 4 predicts: illegal sales replace legal sales one for one until illegal sellers exit. I

Table 3: Elasticities, diversion and tax leakage in 2016: ranges across Models 1–9. Each range is the minimum and maximum across the models for which the measure is defined, not a confidence interval; a range with equal endpoints means that the models give the same value. Store elasticities and margins are medians across stores within each model. Elasticities and store diversion hold other prices, demand shocks and the number of illegal sellers fixed. The all-cannabis elasticity is for a common increase in legal and illegal prices. Tax leakage follows an excise increase, with legal prices re-solved and the number of illegal sellers either fixed or set by zero profit. Sector elasticities aggregate quantity changes over baseline quantities; diversion and leakage divide destination gains by origin losses.

	Nested logit	MNL
Median store own-price elasticity	[−78.48, −19.51]	[−4.10, −1.97]
Legal-sector own-price elasticity	[−4.04, −1.06]	[−3.07, −1.48]
Illegal-sector own-price elasticity	[−2.02, −1.51]	[−1.53, −1.04]
All-cannabis own-price elasticity	[−1.30, −0.98]	[−0.99, −0.67]
Cross-price: illegal quantity response to legal prices	[1.08, 1.44]	[0.74, 1.10]
Cross-price: legal quantity response to illegal price	[1.14, 1.52]	[0.78, 1.16]
Tax leakage (%)	[59.5, 63.9]	[59.4, 100.0]
Store diversion to rival legal stores (%)	[86.3, 95.3]	[25.1, 25.1]
Store diversion to illegal sellers (%)	[2.8, 8.2]	[44.5, 44.5]
Store diversion to outside good (%)	[1.9, 5.6]	[30.4, 30.4]
Median store margin (%)	[1.3, 5.1]	[24.4, 50.8]

focus on nested logit, under which the tax can also change total consumption.

Before turning to taxes, I report price elasticities of demand in 2016 (Table 3; Appendix A.3.7 defines the elasticities and the leakage share). When every legal price rises by 1% and the number of illegal sellers adjusts through zero profit, legal quantity falls by about 3% and total cannabis quantity by about 0.5%. Total demand is inelastic, but legal demand is much more elastic, because buyers can switch to illegal sellers. Of the legal grams lost, about 60% go to illegal sellers and the rest to the outside good. With entry, this leakage share is $S_F/(S_F+\sigma s_0)$ at 2016 shares, and it is the same for a tax as for a price increase (Corollary 2(iii)). Hollenbeck and Uetake (2021), who leave illegal purchases in the outside good, report a legal-sector elasticity of about −1.

Counterfactual tax revenue traces out a Laffer curve. All rates are statutory retail excise rates, on the same base as Washington's 37% rate (Appendix A.3.6). When only the legal market is modeled (the legal-only model, Model 3, Figure 4), a higher excise rate raises revenue, as Hollenbeck and Uetake (2021) find. When the illegal market is included in the estimation, Washington is on the downward-sloping portion of the Laffer curve, and a lower rate raises revenue. In the main specification, 2016 revenue peaks at 20%, below the 37% rate (Figure 13, which compares

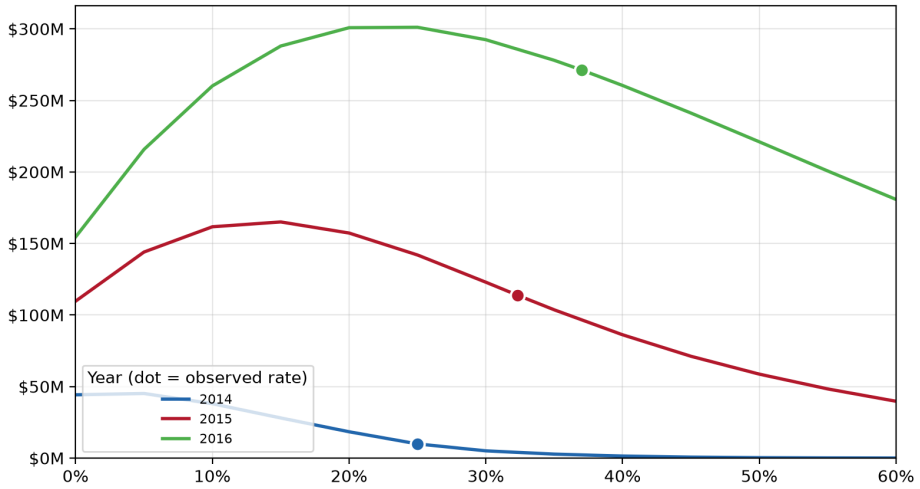


Figure 3: Counterfactual excise revenue as the retail excise rate varies from 0% to 60%, full model. Prices, quantities and the number of illegal sellers are re-solved at each rate; the sales taxes and the licensed set are held fixed.

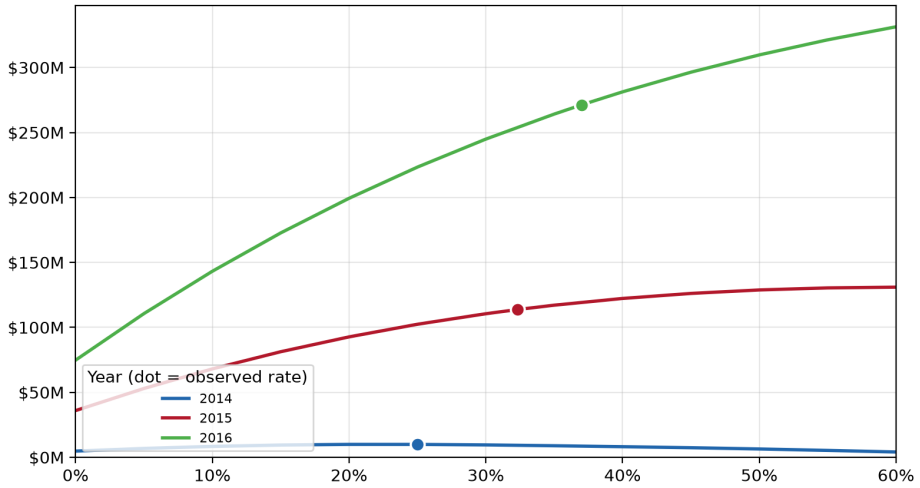


Figure 4: Counterfactual excise revenue as the retail excise rate varies from 0% to 60%, legal-only benchmark, which leaves illegal purchases in the outside good.

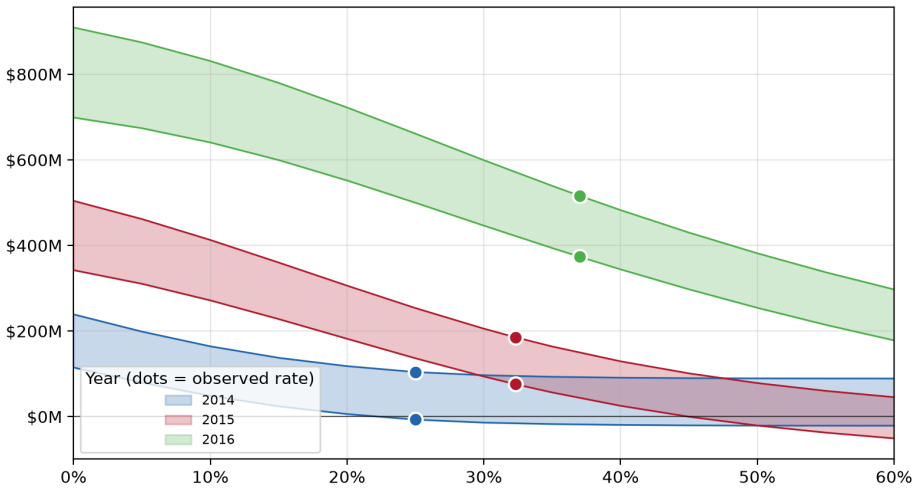


Figure 5: Nested logit: counterfactual total welfare by excise rate, including illegal producer surplus.

the models; Figure 3 shows the curve for Model 6). Figures 14–17 in the Appendix show the counterfactual quantities, illegal entry and legal prices, and Figures 18 and 19 the MNL revenue curves by model. The policymaker chooses the excise rate x to maximize total welfare:

$$\max_{x} \ \text{TW} = \text{CS} + \text{PS} + \text{TR} - C_L\Big(\sum_{\ell} q_{\ell st}\Big) - C_F(q_{Fst}) - C_F^E(N_{st}) \tag{35}$$

When consumer surplus (CS), producer surplus (PS), health costs and externalities are included, the welfare-maximizing rate is lower still. I take the health costs of legal and illegal cannabis from Childs and Stevens (2019) and bound them using their method with data from other sources. I measure the entry externality as the cost of each crime type, from McCollister et al. (2010), times a bounded probability that an entrant commits it. The welfare-maximizing rate depends on which terms the policymaker counts.

A policymaker is unlikely to count illegal profits as welfare, so my main measure excludes illegal PS; Figure 5 includes it for comparison. Both measures are maximized at the lowest rate considered, 0% (Figures 6 and 5). Because cannabis can harm health and productivity, policymakers may also exclude consumer surplus; welfare is then maximized at 15% (Figure 7). Excluding legal profits as well, so that only tax revenue and health and crime costs remain, also gives 15% (Figure 8).

Demand studies of legal cannabis, including Hollenbeck and Uetake (2021), typically put illegal purchases in the outside good. Because the illegal market is large, leaving it out makes legal demand look less elastic in my estimates and moves the 2016 revenue peak above 37%. Unless enforcement eliminates illegal sellers or national legalization lowers legal costs enough for legal products to undercut them, policymakers must set taxes and regulation knowing that legal firms compete with untaxed sellers.

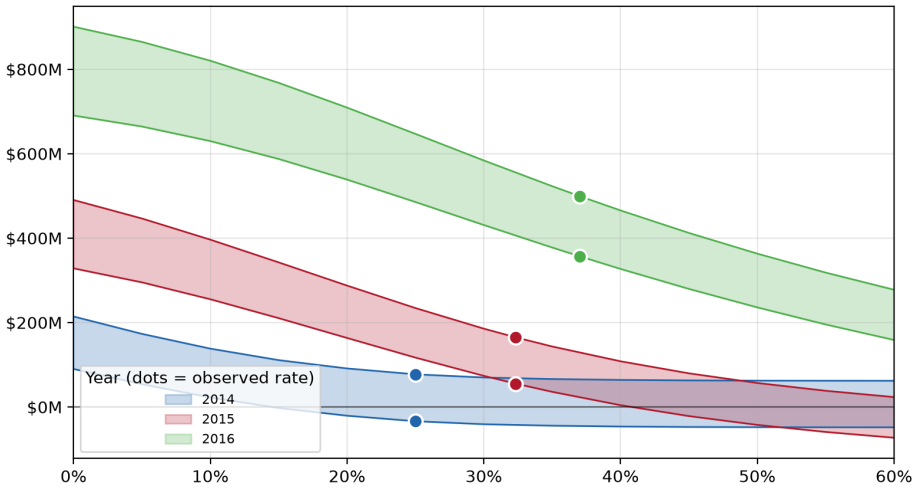


Figure 6: Nested logit: counterfactual total welfare by excise rate, excluding illegal producer surplus (main welfare measure).

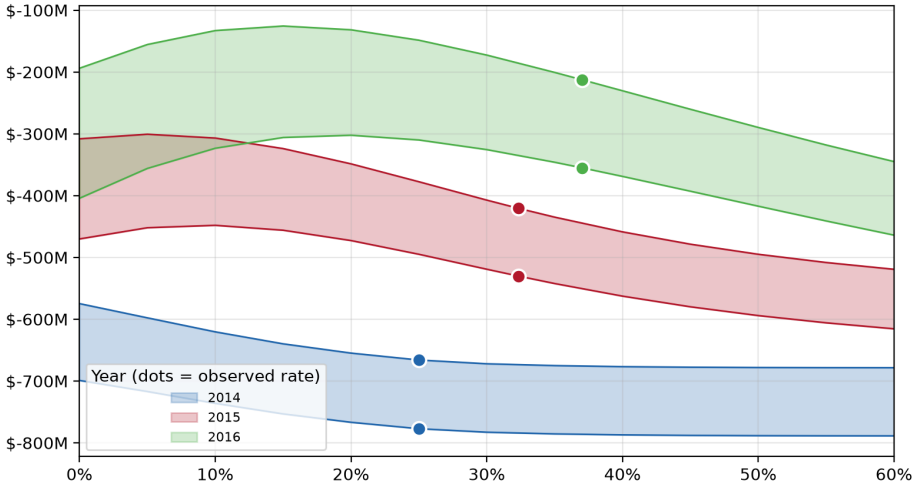


Figure 7: Nested logit: counterfactual total welfare by excise rate, excluding consumer surplus and illegal producer surplus.

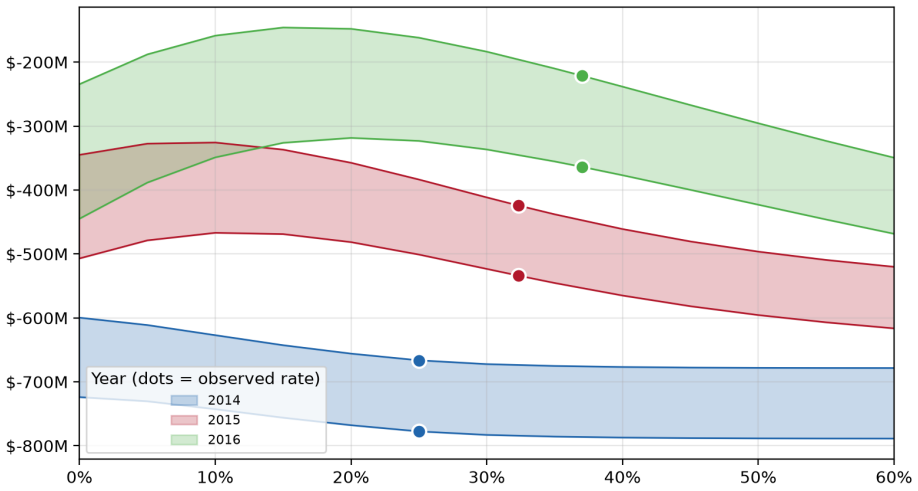


Figure 8: Nested logit: counterfactual total welfare by excise rate, counting only tax revenue and external costs.

I also run two other counterfactuals. In the first, I move illegal quantity into the outside good and rerun the tax counterfactual with the estimated parameters held fixed. This changes the revenue curve little (Figure 20). In the second, I remove the illegal product, re-solve the equilibrium and rerun the tax counterfactual (Figure 21). Without illegal sellers, legal demand is much less elastic, and revenue peaks at a higher rate.

5 Reduced-Form Evidence

The structural model covers one state and three years. To see whether its premise, that legal and illegal sellers compete for the same consumers, holds more widely, I estimate reduced-form models on the state panel. The regressions study how legalization affects illegal prices and the prevalence of past-year and past-month cannabis use. Overall, the illegal price is lower and past-year use is higher after legalization. Past-month use shows no significant effect in the neighbor design, while the larger own-state use estimates come from a design whose event study fails the pre-trend test ($p \leq 0.001$). Because daily and near-daily users account for about 80% of consumption (Caulkins et al., 2020), a rise in past-year use without a rise in past-month use (neighbor design) suggests that legalization adds little consumption. A higher neighbor legal price or tax rate goes with a smaller fall in the illegal price (Tables 4 and 12); the other interactions are insignificant or come from the own-state design. Altogether, these results are consistent with competition between legal and illegal sellers.

I first estimate two-way fixed effects (TWFE) models of recreational sales in a neighboring state or in the own state, interacted with the log legal price (Equations (36) and (37)):

$$y_{st} = \alpha_s + \gamma_t + \beta_1 \cdot O_{st}^{\text{nbr}} + \beta_2 \cdot O_{st}^{\text{nbr}} \times \ln\Big(\min_{n} \; p_{nt}^{\text{rec}}\Big) + \mathbf{X}_{st}'\boldsymbol{\delta} + \varepsilon_{st} \tag{36}$$

$$y_{st} = \alpha_s + \gamma_t + \beta_1 \cdot O_{st}^{\text{own}} + \beta_2 \cdot O_{st}^{\text{own}} \times \ln\big(p_{st}^{\text{rec}}\big) + \mathbf{X}_{st}'\boldsymbol{\delta} + \varepsilon_{st} \tag{37}$$

The outcome y_{st} is the log illegal price, past-year use, or past-month use. O_{st}^{nbr} equals one once any neighboring state has recreational sales, O_{st}^{own} equals one once state s has recreational sales, and p_{nt}^{rec} is the legal retail price in state n; the neighbor design uses the lowest neighboring price. The control vector $\mathbf{X}_{st}$ includes a medical legalization indicator, log population, and the unemployment and poverty rates, and in the neighbor design also an own-state recreational legalization indicator and dispensaries per capita.

Table 4: Two-Way Fixed Effects Estimates: Recreational Sales, Legal Prices, Illegal Prices and Cannabis Use, State Panel 2011–2024

	(1)	(2)	(3)
	Log illegal price	Past-year use	Past-month use
Panel A. Neighbor specification			
Any recreational-sales neighbor	−0.0396***	0.7615**	0.2476
	(0.0123)	(0.3159)	(0.2906)
Neighbor × log min. neighbor price	0.0433***	−0.9421	−0.6002
	(0.0155)	(0.7117)	(0.6438)
Observations / states	659 / 51	710 / 51	710 / 51
Model R^2	0.157	0.311	0.296
Panel B. Own-state specification			
Recreational sales active	−0.0325**	3.3465***	2.7692***
	(0.0147)	(0.4894)	(0.4451)
Recreational sales × log own legal price	0.0301	−2.2662***	−1.4280**
	(0.0219)	(0.6776)	(0.6033)
Observations / states	663 / 51	714 / 51	714 / 51
Model R^2	0.074	0.265	0.237

Notes: Each column reports estimates from a separate regression; Panel A estimates Equation (36) and Panel B Equation (37). The dependent variables are the log illegal retail price (column 1) and National Survey on Drug Use and Health (NSDUH) past-year (column 2) and past-month (column 3) prevalence of cannabis use. Standard errors clustered at the state level are in parentheses. All specifications include a constant, state and year fixed effects, and the controls $\mathbf{X}_{st}$ listed in Section 5. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Next, I estimate event studies of own-state and neighboring-state legalization, without the price interaction (Equation (38)). Here $D_{st}^{k} = \mathbf{1}[t - E_s = k]$, where E_s is the first year with recreational retail sales in the state (own-state design) or in any neighbor (neighbor design); the end bins pool $k \le -4$ and $k \ge 4$, and $k = -1$ is omitted.

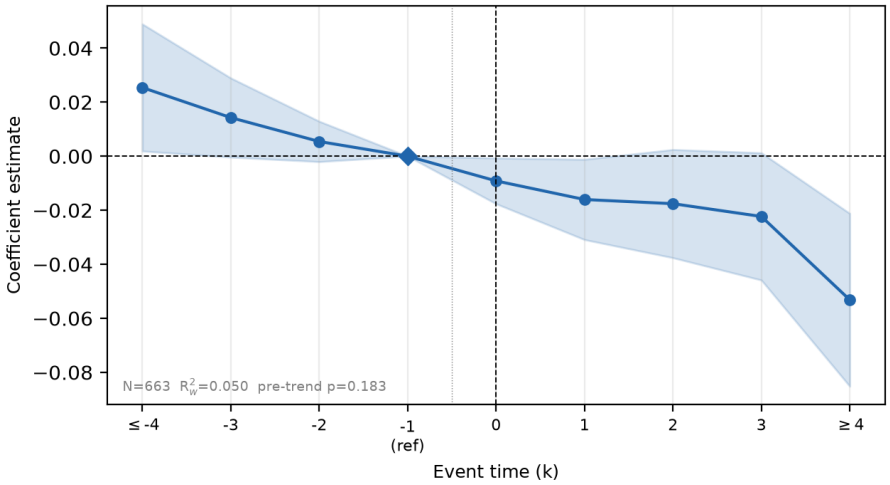


Figure 9: Event Study of Neighboring-State Recreational Sales: Log Illegal Price. Points are the coefficients β_k of Equation (38) with their confidence intervals; event time -1 is the reference year.

$$y_{st} = \alpha_s + \gamma_t + \sum_{k \in \{\leq -4, -3, -2, 0, 1, 2, 3, \geq 4\}} \beta_k \, D^k_{st} + \mathbf{X}'_{st}\boldsymbol{\delta} + \varepsilon_{st} \tag{38}$$

Finally, I replace the price interaction with the log tax rate, $\ln(1+\tau)$ (Equations (39) and (40)).

$$y_{st} = \alpha_s + \gamma_t + \beta_1 \cdot O^{\text{own}}_{st} + \beta_2 \cdot O^{\text{own}}_{st} \times \ln(1+\tau_{st}) + \mathbf{X}'_{st}\boldsymbol{\delta} + \varepsilon_{st} \tag{39}$$

$$y_{st} = \alpha_s + \gamma_t + \beta_1 \cdot O^{\text{nbr}}_{st} + \beta_2 \cdot O^{\text{nbr}}_{st} \times \ln\left(1+\bar{\tau}^{\text{nbr}}_{st}\right) + \mathbf{X}'_{st}\boldsymbol{\delta} + \varepsilon_{st} \tag{40}$$

Here $\bar{\tau}^{\text{nbr}}_{st}$ is the population-weighted average tax rate of the neighboring states.

Past-year use rises by 0.8 percentage points when a neighbor opens, with no significant past-month change; the larger own-state estimates fail the pre-trend test. The illegal price is also lower after legalization. A higher legal price significantly shrinks the response only for the illegal price in the neighbor design (0.043, $p < 0.01$); the other price interactions are insignificant, except those on own-state use, which come from the design that fails the pre-trend test. Most tax interactions are insignificant because own-state tax rates vary little over time. In the neighbor tax model for the illegal price, however, newly legalized neighbors' taxes supply the variation: the illegal price again falls after neighbor legalization, and by less where the neighbors' retail tax rates are higher. Because states open at different dates, two-way fixed effects can use already-treated states as controls (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021); together with the failed own-state pre-trends, this leads me to read the estimates as descriptive. Table 4 and Figures 22–27 report the price-interaction models, Table 11 and Figures 9–11 the event studies (the figures show the neighbor design only), and Table 12 the tax models; all but Table 4 and the event-study figures are

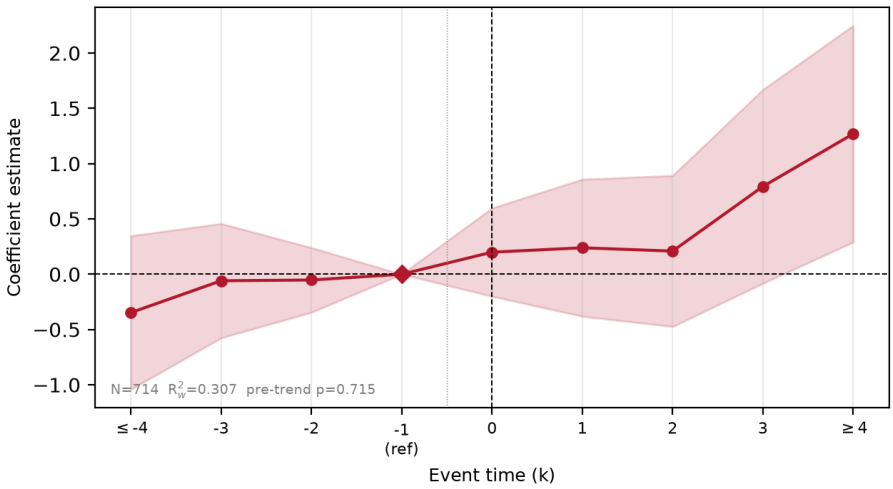


Figure 10: Event Study of Neighboring-State Recreational Sales: Past-Year Use. Points are the coefficients β_k of Equation (38) with their confidence intervals; event time -1 is the reference year.

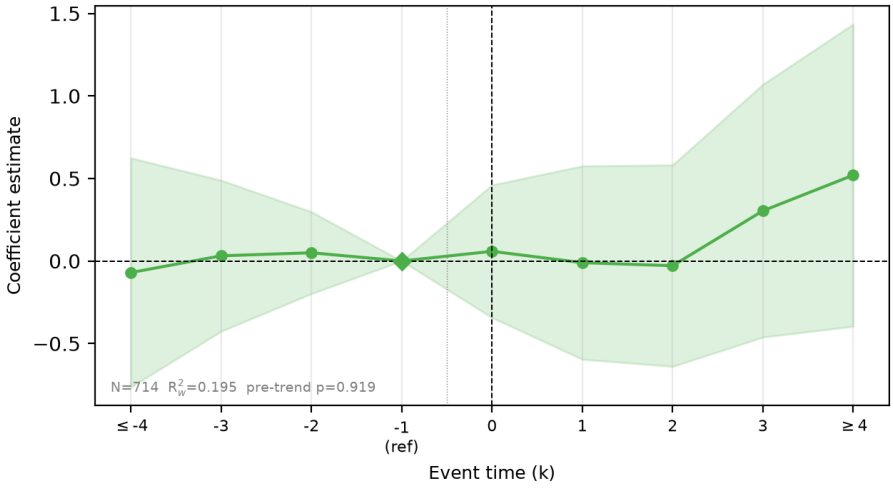


Figure 11: Event Study of Neighboring-State Recreational Sales: Past-Month Use. Points are the coefficients β_k of Equation (38) with their confidence intervals; event time -1 is the reference year.

in the appendix.

6 Conclusion

Higher taxes deter legal cannabis use and, up to the revenue peak, raise revenue. They may also push consumers toward illegal cannabis, which carries more health risks and externalities than legal cannabis through both consumption and production: unregulated products, other criminal activity, and environmental damage around illegal grow sites. Because illegal sellers have cost advantages, illegal cannabis is typically cheaper and captures a large share of potential consumers, even in legal states (Kilmer et al., 2019). In Washington, raising the rate from its baseline shifts consumers to the illegal market: with free entry, about 60% of each gram that legal retailers lose to a tax goes to illegal sellers. Once the illegal market is in the model, the Laffer curve peaks at 20%, below the 37% rate, whereas Hollenbeck and Uetake (2021) conclude that Washington is on the rising side of the Laffer curve. A policymaker who ignores the illegal market therefore sets the rate above the revenue peak, and the state forgoes revenue and incurs these social costs (Figures 13 and 12). I also find that welfare is maximized at rates below the current 37%.

The paper adds to theoretical IO by adapting aggregative games to a dominant oligopoly facing a free-entry fringe, and it takes the resulting model to data from the cannabis industry. Unlike standard demand-and-supply estimation, the empirical model adds a zero-profit condition and arrest indicators, and it provides two ways to recover the number of sellers in a sector where that number is never observed. Estimating illegal entry requires the costs, sales and prices of illegal sellers, which I build from crime reports, surveys, and consulting data. The paper relates to empirical demand estimation, substitution between legal and illegal goods, and the regulation of sin goods and illicit markets.

Reduced-form regressions on the state panel of illegal and legal prices, sales, and cannabis use prevalence show that illegal cannabis prices are lower and past-year use is higher after legalization. Past-month use shows no significant effect in the neighbor design, and the larger own-state use effects come from a design whose event study fails the pre-trend test. These results suggest a small effect on overall use in the neighbor design, coming mainly from less frequent users. In the neighbor design, the illegal price falls by less where the neighbor's legal price or tax is higher. The reduced form cannot count illegal sellers, identify most tax effects (own-state tax rates vary too little), or evaluate counterfactual tax rates, which is why the tax analysis relies on the structural model. The structural model has limits of its own: it covers only Washington over the first three years of legalization, holds the set of licensed retailers fixed, and treats Washington as a closed market although Oregon opened recreational sales in October 2015. Future work could study the spatial sorting of legal and illegal firms across states and municipalities; many municipalities ban

legal cannabis despite state legalization, which fosters illegal entry. Illegal producers in states with legal markets (recreational in California, medical only in Oklahoma) also export to prohibition states, which creates opportunities to study intra-national trade and economic geography.

About 22% of Americans aged 12 and older reported past-year use in 2022 (Substance Abuse and Mental Health Services Administration, 2023), so cannabis regulation has wide-reaching health, environmental, and social implications for users and non-users alike. Regulation already raises the cost of legal production, and excise taxes make it harder still for legal firms to compete, as many consumers switch to illegal cannabis. By giving illegal sellers a cost advantage, taxes and regulation also encourage illegal entry. This mechanism is not specific to Washington. Wherever licensed and unlicensed sellers coexist, a tax on legal cannabis changes how much demand legal and illegal sellers meet, how many illegal sellers the market can support, and, under nested logit, how much is consumed. As with many intoxicating substances, people use cannabis whatever its legal status, so a state could reduce illegal supply by policing it more heavily, which I do not study, or by lowering the tax and regulatory costs that make legal cannabis expensive. The policy lesson is to set the tax with the illegal market in the model. A legal-only model, which leaves illegal purchases in the outside good, places Washington's revenue peak above 37% rather than at 20%, and ignoring the harms of illegal supply sets the tax too high (Section 3.5). Once the illegal fringe is in the model, both the revenue-maximizing rate and, given the external costs I assume, the welfare-maximizing rate lie below Washington's 37%.

References

Acemoglu, D. and M. K. Jensen (2013). Aggregate comparative statics. *Games and Economic Behavior 81*, 27–49.

Adda, J., B. McConnell, and I. Rasul (2014). Crime and the depenalization of cannabis possession: Evidence from a policing experiment. *Journal of Political Economy 122*(5), 1130–1202.

Aguirregabiria, V., A. Iaria, and S. Sokullu (2026). Identification and estimation of demand models with endogenous product entry and exit. SSRN Working Paper 4554232, version of July 27, 2026. An earlier version circulated as CEPR Discussion Paper No. 18398 (2023).

Alfaro, M. and D. Lander (2021). Restricting entry without aggressive pricing. *Research in Economics 75*(4), 305–319.

An, Y., M. R. Baye, Y. Hu, J. Morgan, and M. Shum (2017). Identification and estimation of online price competition with an unknown number of firms. *Journal of Applied Econometrics 32*(1), 80–102.

Anderson, S. P. and Ö. Bedre-Defolie (2024). Hybrid platform model: Monopolistic competition and a dominant firm. *The RAND Journal of Economics 55*(4), 684–718.

Anderson, S. P. and A. de Palma (1992). Multiproduct firms: A nested logit approach. *Journal of Industrial Economics 40*(3), 261–276.

Anderson, S. P. and A. de Palma (2001). Product diversity in asymmetric oligopoly: Is the quality of consumer goods too low? *Journal of Industrial Economics 49*(2), 113–135.

Anderson, S. P. and A. de Palma (2006). Market performance with multiproduct firms. *Journal of Industrial Economics 54*(1), 95–124.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Anderson, S. P., N. Erkal, and D. Piccinin (2014). Aggregative oligopoly games with entry. Draft, January 2014. An earlier version circulated as CEPR Discussion Paper No. 9511 (June 2013).

Anderson, S. P., N. Erkal, and D. Piccinin (2015). Aggregative oligopoly games with entry. Draft, June 2015.

Anderson, S. P., N. Erkal, and D. Piccinin (2020). Aggregative games and oligopoly theory: short-run and long-run analysis. *The RAND Journal of Economics 51*(2), 470–495.

Association of Washington Cities (2021). Cannabis revenues and cities. Slides presented to the Social Equity in Cannabis Task Force, October 19, 2021.

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

Aussel, D. and A. Svensson (2018). Some remarks about existence of equilibria, and the validity of the EPCC reformulation for multi-leader-follower games. *Journal of Nonlinear and Convex Analysis 19*(7), 1141–1162.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology 28*, 97–104.

Bauer, S., J. Olson, A. Cockrill, M. van Hattem, L. Miller, M. Tauzer, and G. Leppig (2015). Impacts of surface water diversions for marijuana cultivation on aquatic habitat in four northwestern California watersheds. *PLoS ONE 10*(3), e0120016.

Becker, G. S., K. M. Murphy, and M. Grossman (2006). The market for illegal goods: The case of drugs. *Journal of Political Economy 114*(1), 38–60.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2), 242–262.

Berry, S. T. and J. Waldfogel (1999). Free entry and social inefficiency in radio broadcasting. *RAND Journal of Economics 30*(3), 397–420.

Bertomeu, J. (2005). Undesirable goods, illegal trade and taxation. Working paper, Tepper School of Business, Carnegie Mellon University, 18 October 2005. EconWPA Industrial Organization 0510005.

Blood, M. and C. Weber (2024, February). 6 killings in likely cannabis dispute spotlight violence in California's illegal pot market. PBS NewsHour (Associated Press).

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika 61*(1), 109–121.

Breusch, T. (2005). Estimating the underground economy using MIMIC models. Econometrics Working Paper 0507003, EconWPA. Posted July 12, 2005; revised December 15, 2005. Working paper, not peer reviewed.

California Department of Tax and Fee Administration (2025). Cannabis retailers.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Carrieri, V., L. Madio, and F. Principe (2019). Light cannabis and organized crime: Evidence from (unintended) liberalization in Italy. *European Economic Review 113*, 63–76.

Caulkins, J. P., B. Pardo, and B. Kilmer (2020). Intensity of cannabis use: Findings from three online surveys. *International Journal of Drug Policy 79*, 102740.

Childs, J. and J. Stevens (2019). The state must compete: Optimal pricing of legal cannabis. *Canadian Public Administration 62*(4), 656–673.

Ciliberto, F., C. Murry, and E. Tamer (2021). Market structure and competition in airline markets. *Journal of Political Economy 129*(11), 2995–3038.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with PyBLP. *The RAND Journal of Economics 51*(4), 1108–1161.

Conlon, C. and N. L. Rao (2023). The cost of curbing externalities with market power: Alcohol regulations and tax alternatives. NBER Working Paper 30896, National Bureau of Economic Research.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

Corchón, L. C. (1994). Comparative statics for aggregative games: The strong concavity case. *Mathematical Social Sciences 28*(3), 151–165.

Davis, A. J., K. R. Geisler, and M. W. Nichols (2016). The price elasticity of marijuana demand: Evidence from crowd-sourced transaction data. *Empirical Economics 50*(4), 1171–1192.

Davis, C., M. E. Hill, and R. Phillips (2019, January). Taxing cannabis. Technical report, Institute on Taxation and Economic Policy.

De Loecker, J. and P. T. Scott (2026). Markup estimation using production and demand data: An application to the U.S. brewing industry. *Review of Economic Studies*. Advance article.

DeCicca, P., D. Kenkel, and F. Liu (2013). Excise tax avoidance: The case of state cigarette taxes. *Journal of Health Economics 32*(6), 1130–1141.

Dills, A. K., S. Goffard, and J. Miron (2017, September). The effects of marijuana liberalizations: Evidence from Monitoring the Future. NBER Working Paper 23779, National Bureau of Economic Research.

Dragone, D., G. Prarolo, P. Vanin, and G. Zanella (2019). Crime and the legalization of recreational marijuana. *Journal of Economic Behavior & Organization 159*, 488–501.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*(5), 2231–2267.

Dubois, P., R. Griffith, and M. O'Connell (2020). How well targeted are soda taxes? *American Economic Review 110*(11), 3661–3704.

Etro, F. (2006). Aggressive leaders. *The RAND Journal of Economics 37*(1), 146–154.

Etro, F. (2008). Stackelberg competition with endogenous entry. *The Economic Journal 118*(532), 1670–1697.

Etro, F. (2023). Hybrid marketplaces with free entry of sellers. *Review of Industrial Organization 62*(2), 119–148.

Farronato, C. and A. Fradkin (2022). The welfare effects of peer entry: The case of Airbnb and the accommodation industry. *American Economic Review 112*(6), 1782–1817.

Forchheimer, K. (1908). Theoretisches zum unvollständigen Monopole. *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich 32*(1), 1–12.

Frey, B. S. and H. Weck-Hannemann (1984). The hidden economy as an 'unobserved' variable. *European Economic Review 26*(1–2), 33–53.

Fryer, Jr., R. G., P. S. Heaton, S. D. Levitt, and K. M. Murphy (2013). Measuring crack cocaine and its impact. *Economic Inquiry 51*(3), 1651–1681.

Galenianos, M. and A. Gavazza (2017). A structural model of the retail market for illicit drugs. *American Economic Review 107*(3), 858–896.

Galenianos, M., R. L. Pacula, and N. Persico (2012). A search-theoretic model of the retail market for illicit drugs. *Review of Economic Studies 79*(3), 1239–1269.

Gandhi, A. and J.-F. Houde (2019). Measuring substitution patterns in differentiated-products industries. NBER Working Paper 26375, National Bureau of Economic Research.

Gavrilova, E., T. Kamada, and F. Zoutman (2019, January). Is legal pot crippling Mexican drug trafficking organisations? The effect of medical marijuana laws on US crime. *The Economic Journal 129*(617), 375–407.

Golombek, R., A. A. Irarrazabal, and L. Ma (2018). OPEC's market power: An empirical dominant firm model for the oil market. *Energy Economics 70*, 98–115.

Golub, G. H. and V. Pereyra (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis 10*(2), 413–432.

Goodman, S., E. Wadsworth, and D. Hammond (2022). Reasons for purchasing cannabis from illegal sources in legal markets: Findings among cannabis consumers in Canada and U.S. states, 2019–2020. *Journal of Studies on Alcohol and Drugs 83*(3), 392–401.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Han, J., K. A. Mottaleb, J. N. Ng'ombe, and A. Durand-Morat (2024). Does sin tax on the legal market facilitate the illicit market? An ex-ante assessment on the US cannabis market. *Journal of the Agricultural and Applied Economics Association 3*(2), 330–345.

Hansen, B., K. Houghton, K. Miller, and C. Weber (2026). Shifting tax incidence: Evidence from the Washington State cannabis market. *Journal of Policy Analysis and Management 45*(3), e70041. Earlier version: NBER Working Paper 23632 (2017), "Getting into the Weeds of Tax Invariance".

Hansen, B., K. Miller, B. Seo, and C. Weber (2020, June). Taxing the potency of sin goods: Evidence from recreational cannabis and liquor markets. *National Tax Journal 73*(2), 511–544.

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Hao, Z. and B. W. Cowan (2020). The cross-border spillover effects of recreational marijuana legalization. *Economic Inquiry 58*(2), 642–666.

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

Hu, M. and M. Fukushima (2015). Multi-leader-follower games: models, methods and applications. *Journal of the Operations Research Society of Japan 58*(1), 1–23.

Ino, H. and T. Matsumura (2012). How many firms should be leaders? Beneficial concentration revisited. *International Economic Review 53*(4), 1323–1340.

Irvine, I. and M. Light (2020). The tax consequences of legal cannabis. *Canadian Public Policy 46*(3), 305–322.

Jacobi, L. and M. Sovinsky (2016). Marijuana on Main Street? Estimating demand in markets with limited access. *American Economic Review 106*(8), 2009–2045.

Jöreskog, K. G. and A. S. Goldberger (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Kahai, S. K., D. L. Kaserman, and J. W. Mayo (1996). Is the "dominant firm" dominant? An empirical analysis of AT&T's market power. *Journal of Law and Economics 39*(2), 499–517.

Kaste, M. (2024, April). Black market cannabis thrives in California despite legalization. NPR.

Kerr, W. C. and Y. Ye (2022). Estimating usual grams per day of marijuana use from purchases. *Addiction Research & Theory 30*(5), 360–367.

Kilmer, B. (2019, August). Americans' spending on illicit drugs nears $150 billion annually; appears to rival what is spent on alcohol. RAND Corporation news release.

Kilmer, B., S. Davenport, R. Smart, J. P. Caulkins, and G. Midgette (2019). After the grand opening: Assessing cannabis supply and demand in Washington State. Research Report RR-3138-WSLCB, RAND Corporation.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Mace, C., E. Patel, and N. Seegert (2020). Marijuana taxation and imperfect competition. *National Tax Journal 73*(2), 545–592.

MacKay, A. and N. H. Miller (2025). Estimating models of supply and demand: Instruments and covariance restrictions. *American Economic Journal: Microeconomics 17*(1), 238–281.

Mallozzi, L. and R. Messalli (2017). Multi-leader multi-follower model with aggregative uncertainty. *Games 8*(3), 25.

Mann, B. (2025, February). U.S. cannabis shoppers face a market flush with illegal weed. NPR.

McCollister, K. E., M. T. French, and H. Fang (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence 108*(1), 98–109.

Meinhofer, A. and A. Rubli (2021). Illegal drug market responses to state recreational cannabis laws. *Addiction 116*(12), 3433–3443.

Miller, A. M., R. Rosenman, and B. W. Cowan (2017). Recreational marijuana legalization and college student use: Early evidence. *SSM - Population Health 3*, 649–657.

Miller, K. and B. Seo (2021, March). The effect of cannabis legalization on substance demand and tax revenues. *National Tax Journal 74*(1), 107–145.

Miller, N. H. and M. C. Weinberg (2017). Understanding the price effects of the MillerCoors joint venture. *Econometrica 85*(6), 1763–1791.

Miravete, E. J., K. Seim, and J. Thurk (2018). Market power and the Laffer curve. *Econometrica 86*(5), 1651–1687.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics 3*(4), 370–379.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

New Frontier Data (2023). Cannabis in America: 2023 U.S. cannabis report. Technical report, New Frontier Data. Reports $30 billion in state-legal sales and an estimated $76 billion in illicit sales for 2022.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: Elsevier.

Nocke, V. and N. Schutz (2018). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica 86*(2), 523–557.

Parenti, M. (2018). Large and small firms in a global market: David vs. Goliath. *Journal of International Economics 110*, 103–118.

Perrault, T. (2023). Flying high? Legalization and the black market for cannabis. Job market paper, McGill University, November 12, 2023.

Plúas-López, A. D. (2024). *Three Essays on Applied Microeconomics*. Ph. D. thesis, Universität Mannheim. Chapter 1 studies legal and drug-trafficking sources of marijuana in Uruguay; earlier version: job market paper, "Beyond Legalization: Access and Use of (Non-)Drug-Trafficking Marijuana" (2023).

Rojas, C. and A. Briceño (2019). The effects of piracy on competition: Evidence from subscription TV. *International Journal of Industrial Organization 63*, 18–43.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*(5), 999–1033.

Schneider, F. and D. H. Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature 38*(1), 77–114.

Selten, R. (1970). *Preispolitik der Mehrproduktenunternehmung in der Statischen Theorie*. Ökonometrie und Unternehmensforschung. Berlin, Germany: Springer-Verlag.

Shimomura, K.-I. and J.-F. Thisse (2012). Competition among the big and the small. *RAND Journal of Economics 43*(2), 329–347.

Smith, R. T. (1976). The legal and illegal markets for taxed goods: Pure theory and an application to state government taxation of distilled spirits. *Journal of Law and Economics 19*(2), 393–429.

Sovinsky, M., L. Jacobi, A. Allocca, and T. Sun (2024). More than joints: Multi-substance use, choice limitations, and policy implications. CRC TR 224 Discussion Paper 501, University of Bonn and University of Mannheim. CRC TR 224 Discussion Paper 501 (February 2024); also CEPR Discussion Paper 18693 (December 2023).

Substance Abuse and Mental Health Services Administration (2023). Key substance use and mental health indicators in the United States: Results from the 2022 National Survey on Drug Use and Health. HHS Publication PEP23-07-01-006, Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.

Thomas, D. (2019). License quotas and the inefficient regulation of sin goods: Evidence from the Washington recreational marijuana market. SSRN Scholarly Paper 3312960, Social Science Research Network.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review 108*(8), 2015–2047.

White, A. P. (1981). The dominant firm structure: Theoretical myth or empirical reality? *Southern Economic Journal 48*(2), 427–436.

Zamarra, L. (2013, August). Modeling marijuana businesses and costs of legal compliance. I-502 project #430-5b, final report, BOTEC Analysis Corp. August 10, 2013. Published by the Washington State Liquor and Cannabis Board.

A Appendix

A.0 Tables & Figures

A.0.1 Structural Model Estimates & Counterfactuals

Table 2 in Section 4.4 lists the nine models and their numbers.

Table 5: Nested logit estimates for Washington counties, 2014–2016: benchmarks without entry (Models 1–3)

Parameter	Demand only Model 1	No entry Model 2	Legal only Model 3
Demand			
Available brands	0.295 (0.116)	0.214 (0.130)	0.253 (0.074)
Flower share	−5.492 (0.756)	−8.192 (0.934)	−1.461 (0.396)
Illegal-market indicator	−0.693 (0.205)	−0.513 (0.216)	—
Price coefficient (α)	0.377 (0.023)	0.330 (0.024)	0.115 (0.008)
Nesting parameter (σ)	0.943 (0.044)	0.936 (0.049)	0.932 (0.031)
Supply			
Wholesale price	—	0.869 (0.064)	0.439 (0.080)
Available brands	—	0.044 (0.020)	−0.005 (0.017)
Flower share	—	0.070 (0.127)	−0.037 (0.011)
Illegal-market intercept	—	−0.112 (0.015)	—

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Model 1 has no supply or entry equation, Models 2 and 3 have no entry equation, and Model 3 uses legal products only, so it has no illegal-market indicator.

Table 6: Nested logit estimates for Washington counties, 2014–2016: substitution route, zero profit substituted for entry (Models 4–6)

Parameter	Substitution, constant entry cost Model 4	Substitution, entry-cost shifters Model 5	Substitution with arrests Model 6
Demand			
Available brands	0.477 (0.097)	0.346 (0.136)	0.342 (0.129)
Flower share	−5.768 (0.774)	−7.425 (0.804)	−6.778 (0.855)
Illegal-market indicator	−0.866 (0.136)	−0.674 (0.159)	−1.065 (0.181)
Price coefficient (α)	0.398 (0.023)	0.378 (0.024)	0.440 (0.024)
Nesting parameter (σ)	0.915 (0.022)	0.901 (0.053)	0.935 (0.045)
Supply			
Wholesale price	0.859 (0.044)	0.950 (0.047)	0.901 (0.046)
Available brands	0.009 (0.012)	0.021 (0.013)	−0.001 (0.013)
Flower share	0.258 (0.107)	0.689 (0.131)	0.013 (0.007)
Illegal-market intercept	−0.080 (0.012)	−0.068 (0.013)	−0.061 (0.013)
Entry-cost shifters			
Officers (log)	—	−1.699 (1.322)	−2.281 (2.146)
Lagged property incidents (log)	—	0.498 (0.815)	1.373 (1.520)
Housing value (log Zillow Home Value Index)	—	2.006 (2.682)	0.879 (1.247)
Electricity price (log)	—	−2.632 (1.809)	−0.323 (1.403)
Commercial value per acre (log)	—	−0.688 (0.843)	−0.990 (0.963)
Urban share	—	7.462 (4.386)	6.817 (5.002)
Urbanized-area share	—	10.793 (5.979)	6.302 (4.680)
Population density (log)	—	−0.014 (0.575)	1.441 (0.969)
Indicator entry loadings			
Sales and manufacturing arrests, loading (λ_s)	—	—	1.745 (7.590)
Possession arrests, loading (λ_p)	—	—	0.681 (2.917)
Annual entry			
N_{2014}	258.5	472.0	316.7
N_{2015}	195.9	368.6	238.9
N_{2016}	160.1	287.8	189.1

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Annual entry N_t is the recovered number of illegal sellers in year t, summed over counties.

Table 7: Nested logit estimates for Washington counties, 2014–2016: concentration route, entry recovered from several readings (Models 7–9)

Parameter	Two readings, demand and zero profit Model 7	Four readings, fixed weights Model 8	Four readings, covariance weights Model 9
Demand			
Available brands	0.513 (0.149)	0.511 (0.098)	0.486 (0.132)
Flower share	−5.213 (0.823)	−5.276 (0.737)	−5.917 (0.781)
Illegal-market indicator	−0.548 (0.136)	−0.469 (0.115)	−0.573 (0.142)
Price coefficient (α)	0.342 (0.026)	0.335 (0.022)	0.362 (0.024)
Nesting parameter (σ)	0.830 (0.055)	0.830 (0.001)	0.850 (0.047)
Supply			
Wholesale price	1.189 (0.080)	1.131 (0.077)	1.097 (0.075)
Available brands	0.059 (0.021)	0.052 (0.020)	0.051 (0.020)
Flower share	0.193 (0.021)	0.186 (0.020)	0.182 (0.020)
Illegal-market intercept	−0.121 (0.017)	−0.114 (0.015)	−0.111 (0.016)
Entry-cost shifters			
Officers (log)	−0.216 (0.318)	−0.059 (0.169)	0.044 (0.250)
Lagged property incidents (log)	0.264 (0.475)	0.754 (0.370)	0.350 (0.377)
Housing value (log Zillow Home Value Index)	0.496 (1.191)	1.014 (0.390)	0.479 (0.247)
Electricity price (log)	−2.082 (1.025)	0.260 (0.198)	0.040 (0.149)
Commercial value per acre (log)	−0.447 (0.421)	−0.028 (0.048)	−0.017 (0.035)
Urban share	2.109 (1.400)	1.174 (0.244)	0.634 (0.229)
Urbanized-area share	3.784 (1.726)	0.814 (0.223)	0.429 (0.177)
Population density (log)	0.054 (0.351)	0.104 (0.049)	0.157 (0.031)
Indicator entry loadings			
Sales and manufacturing arrests, loading (λ_s)	—	−1.528 (0.868)	−6.701 (17.578)
Possession arrests, loading (λ_p)	—	−0.760 (0.461)	−3.216 (8.627)
Direct controls C_{st} in the sales and manufacturing arrest equation (ψ_s)			
Other-drug incidents (log)	—	-0.128^{c} (0.243)	-1.339^{c} (3.672)
Officers (log)	—	1.074^{c} (0.633)	5.095^{c} (13.094)
Direct controls C_{st} in the possession arrest equation (ψ_p)			
Other-drug incidents (log)	—	0.345^{c} (0.163)	-0.216^{c} (1.840)
Officers (log)	—	0.497^{c} (0.343)	2.407^{c} (6.437)
Annual entry			
N_{2014}	422.3	423.6	281.7
N_{2015}	287.2	256.1	203.5
N_{2016}	205.6	169.5	156.1

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Annual entry N_t is the recovered number of illegal sellers in year t, summed over counties. ^{c} Coefficient on a direct control in an arrest equation; its standard error is conditional on the recovered entry.

Table 8: Multinomial logit (MNL, $\sigma = 0$) estimates for Washington counties, 2014–2016: benchmarks without entry (Models 1–3)

Parameter	Demand only Model 1	No entry Model 2	Legal only Model 3
Demand			
Available brands	2.141 (0.143)	1.993 (0.151)	2.188 (0.138)
Flower share	−6.373 (1.807)	−7.261 (1.976)	−3.863 (2.023)
Illegal-market indicator	2.175 (0.251)	1.626 (0.262)	—
Price coefficient (α)	0.227 (0.031)	0.319 (0.035)	0.161 (0.053)
Nesting parameter (σ)	0 (fixed)	0 (fixed)	0 (fixed)
Supply			
Wholesale price	—	1.191 (0.081)	1.148 (0.449)
Available brands	—	0.026 (0.020)	0.002 (0.025)
Flower share	—	0.688 (0.176)	−0.013 (0.015)
Illegal-market intercept	—	−0.277 (0.035)	—

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Model 1 has no supply or entry equation, Models 2 and 3 have no entry equation, and Model 3 uses legal products only, so it has no illegal-market indicator.

Table 9: Multinomial logit (MNL, $\sigma = 0$) estimates for Washington counties, 2014–2016: substitution route, zero profit substituted for entry (Models 4–6)

Parameter	Substitution, constant entry cost Model 4	Substitution, entry-cost shifters Model 5	Substitution with arrests Model 6
Demand			
Available brands	1.957 (0.156)	2.205 (0.147)	2.291 (0.142)
Flower share	−1.643 (1.896)	−8.747 (1.911)	−3.189 (1.748)
Illegal-market indicator	−1.572 (0.185)	−1.677 (0.228)	−1.357 (0.160)
Price coefficient (α)	0.274 (0.031)	0.335 (0.039)	0.284 (0.029)
Nesting parameter (σ)	0 (fixed)	0 (fixed)	0 (fixed)
Supply			
Wholesale price	1.197 (0.084)	1.233 (0.074)	1.163 (0.081)
Available brands	0.027 (0.021)	0.024 (0.016)	0.012 (0.018)
Flower share	0.449 (0.162)	0.774 (0.162)	0.023 (0.010)
Illegal-market intercept	−0.348 (0.048)	−0.246 (0.035)	−0.315 (0.041)
Entry-cost shifters			
Officers (log)	—	−0.512 (0.282)	0.008 (0.147)
Lagged property incidents (log)	—	0.485 (0.264)	−0.127 (0.161)
Housing value (log Zillow Home Value Index)	—	−0.759 (0.538)	−0.981 (0.389)
Electricity price (log)	—	0.316 (0.675)	−1.514 (0.327)
Commercial value per acre (log)	—	−0.296 (0.173)	−0.103 (0.146)
Urban share	—	7.378 (1.350)	4.211 (0.608)
Urbanized-area share	—	7.262 (1.305)	4.426 (0.716)
Population density (log)	—	−0.110 (0.158)	0.145 (0.126)
Indicator entry loadings			
Sales and manufacturing arrests, loading (λ_s)	—	—	−0.387 (0.074)
Possession arrests, loading (λ_p)	—	—	−0.121 (0.062)
Annual entry			
N_{2014}	4,439.2	2,844.4	2,016.0
N_{2015}	3,365.0	2,251.5	1,732.9
N_{2016}	2,748.8	2,002.7	1,619.7

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Annual entry N_t is the recovered number of illegal sellers in year t, summed over counties.

Table 10: Multinomial logit (MNL, $\sigma = 0$) estimates for Washington counties, 2014–2016: concentration route, entry recovered from several readings (Models 7–9)

Parameter	Two readings, demand and zero profit Model 7	Four readings, fixed weights Model 8	Four readings, covariance weights Model 9
Demand			
Available brands	2.883 (0.165)	2.745 (0.159)	2.843 (0.164)
Flower share	−1.795 (1.400)	−0.952 (1.328)	−1.599 (1.389)
Illegal-market indicator	−1.352 (0.135)	−1.155 (0.145)	−1.207 (0.130)
Price coefficient (α)	0.252 (0.026)	0.229 (0.025)	0.248 (0.026)
Nesting parameter (σ)	0 (fixed)	0 (fixed)	0 (fixed)
Supply			
Wholesale price	1.275 (0.094)	1.319 (0.103)	1.258 (0.096)
Available brands	−0.037 (0.021)	−0.024 (0.022)	−0.033 (0.021)
Flower share	0.026 (0.010)	0.025 (0.011)	0.021 (0.010)
Illegal-market intercept	−0.388 (0.054)	−0.440 (0.070)	−0.404 (0.057)
Entry-cost shifters			
Officers (log)	0.128 (0.076)	−0.331 (0.159)	0.203 (0.096)
Lagged property incidents (log)	−0.062 (0.141)	0.008 (0.190)	−0.118 (0.144)
Housing value (log Zillow Home Value Index)	−0.499 (0.377)	−1.775 (0.480)	−1.332 (0.296)
Electricity price (log)	−1.496 (0.231)	−1.649 (0.438)	−1.380 (0.207)
Commercial value per acre (log)	−0.270 (0.156)	0.052 (0.172)	0.121 (0.143)
Urban share	2.775 (0.457)	5.309 (0.681)	2.957 (0.292)
Urbanized-area share	3.411 (0.506)	5.506 (0.768)	2.946 (0.392)
Population density (log)	0.099 (0.113)	0.239 (0.133)	0.184 (0.109)
Indicator entry loadings			
Sales and manufacturing arrests, loading (λ_s)	—	0.105 (0.133)	13.860 (20.172)
Possession arrests, loading (λ_p)	—	0.489 (0.124)	7.765 (11.026)
Direct controls C_{st} in the sales and manufacturing arrest equation (ψ_s)			
Other-drug incidents (log)	—	1.017[c] (0.081)	2.155[c] (1.717)
Officers (log)	—	−0.083[c] (0.029)	−0.982[c] (1.326)
Direct controls C_{st} in the possession arrest equation (ψ_p)			
Other-drug incidents (log)	—	0.980[c] (0.035)	1.553[c] (0.944)
Officers (log)	—	−0.107[c] (0.018)	−0.583[c] (0.725)
Annual entry			
N_{2014}	2,486.8	2,788.0	2,243.2
N_{2015}	1,555.6	1,714.5	1,393.5
N_{2016}	1,380.7	1,418.7	1,227.8

Notes: GMM estimates with standard errors in parentheses; — denotes not estimated or not applicable. Annual entry N_t is the recovered number of illegal sellers in year t, summed over counties. [c] Coefficient on a direct control in an arrest equation; its standard error is conditional on the recovered entry.

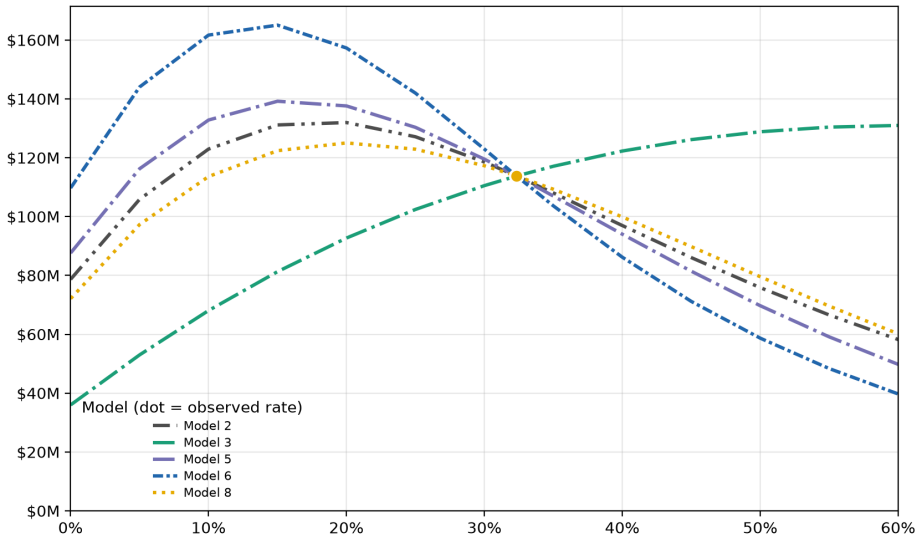


Figure 12: Nested logit: counterfactual tax revenue by excise rate in 2015, one curve per model, with prices, quantities and illegal entry solved together at each rate. Legend numbers are the model numbers in Table 2.

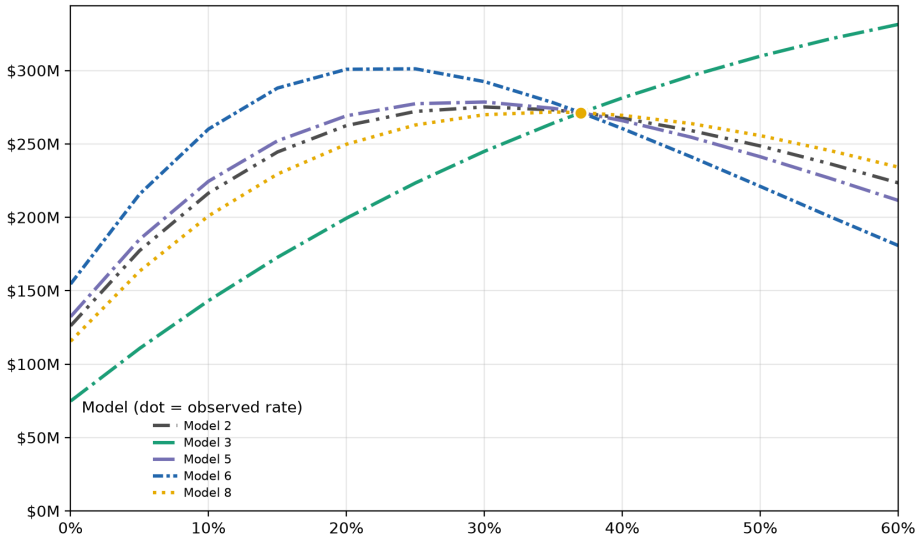


Figure 13: Nested logit: counterfactual tax revenue by excise rate in 2016, one curve per model, with prices, quantities and illegal entry solved together at each rate. Legend numbers are the model numbers in Table 2.

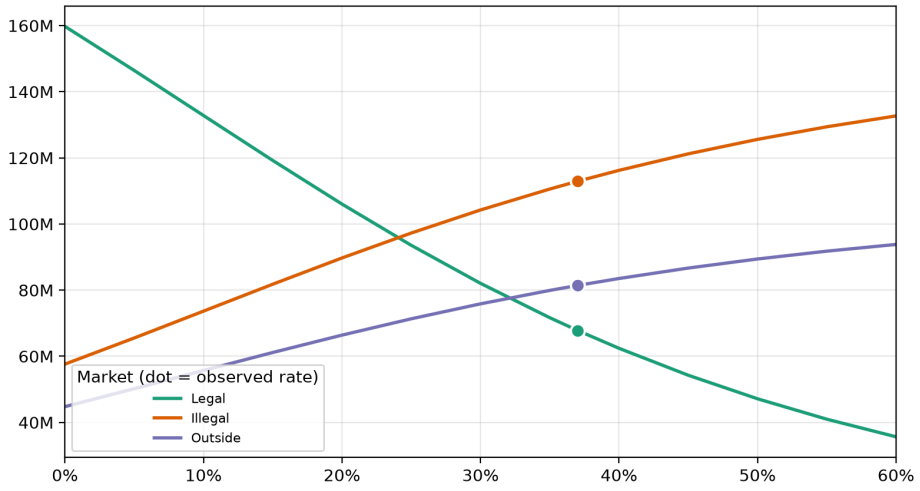


Figure 14: Nested logit, Model 6: counterfactual legal, illegal and outside-good quantities by excise rate, 2016

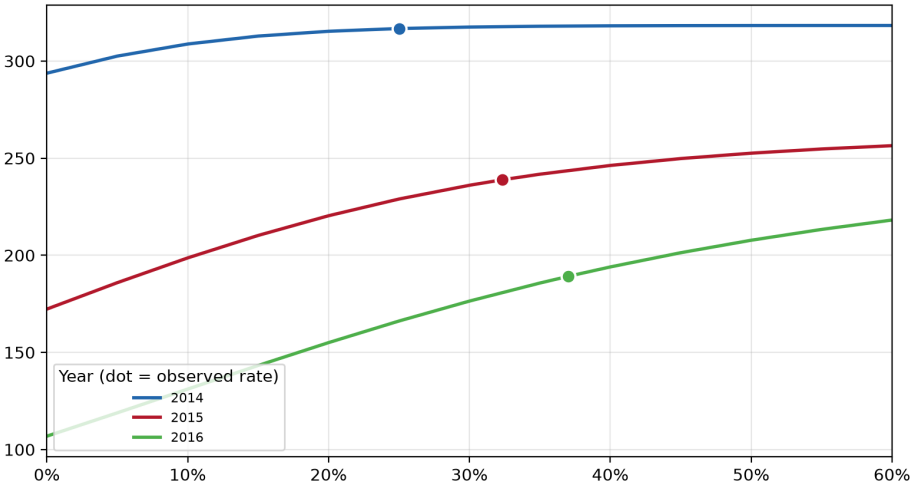


Figure 15: Nested logit, Model 6: counterfactual number of illegal sellers by excise rate and year

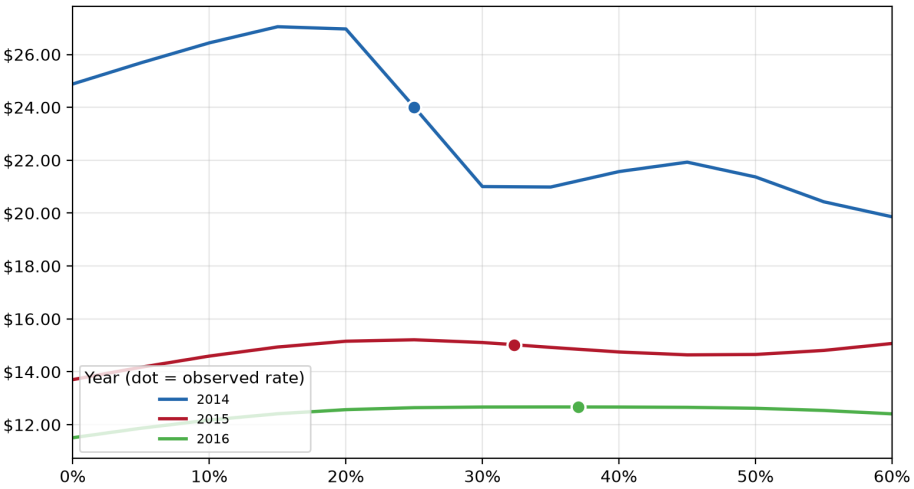


Figure 16: Nested logit, Model 6 (substitution with arrests): counterfactual legal prices by excise rate

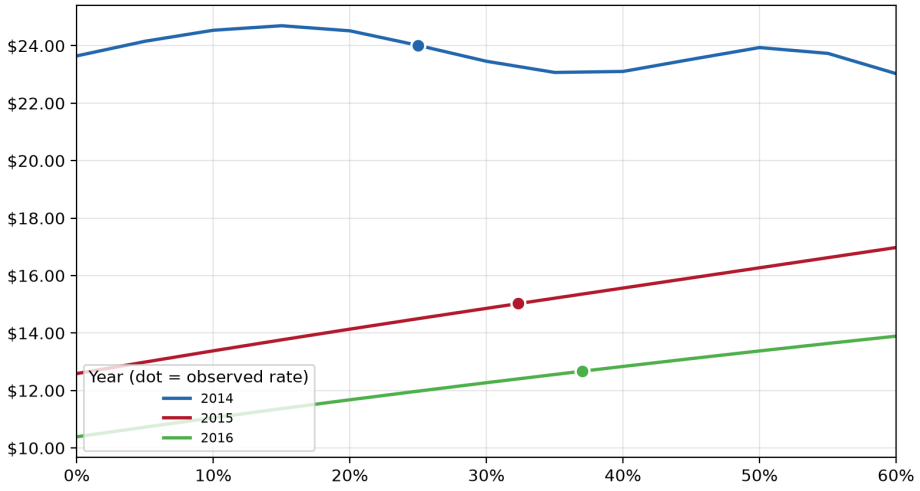


Figure 17: Nested logit, Model 8 (four readings, fixed weights): counterfactual legal prices by excise rate

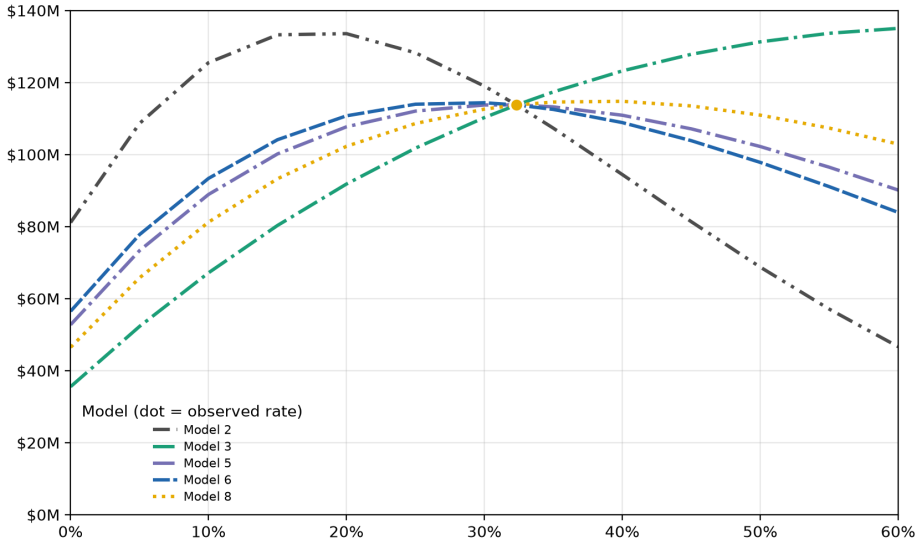


Figure 18: Multinomial logit (MNL): counterfactual tax revenue by excise rate in 2015, one curve per model. Legend numbers are the model numbers in Table 2.

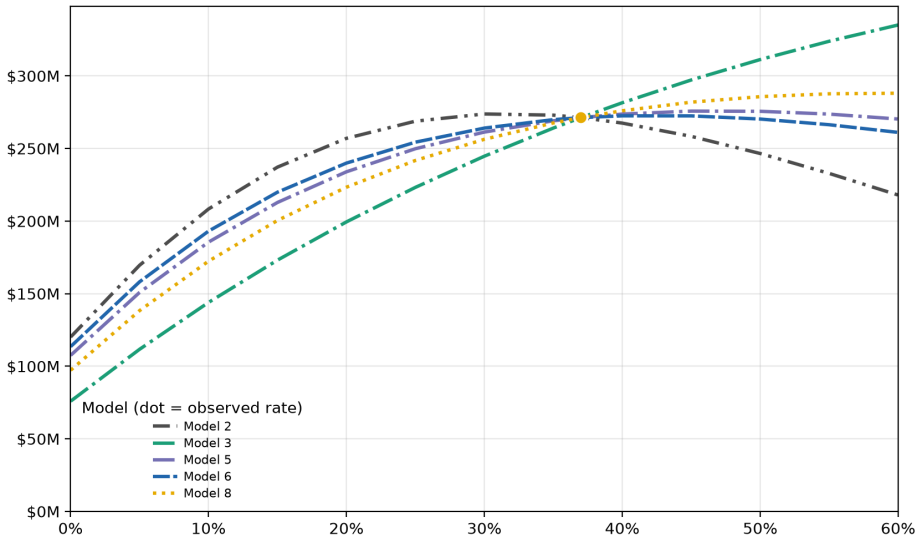


Figure 19: Multinomial logit (MNL): counterfactual tax revenue by excise rate in 2016, one curve per model. Legend numbers are the model numbers in Table 2.

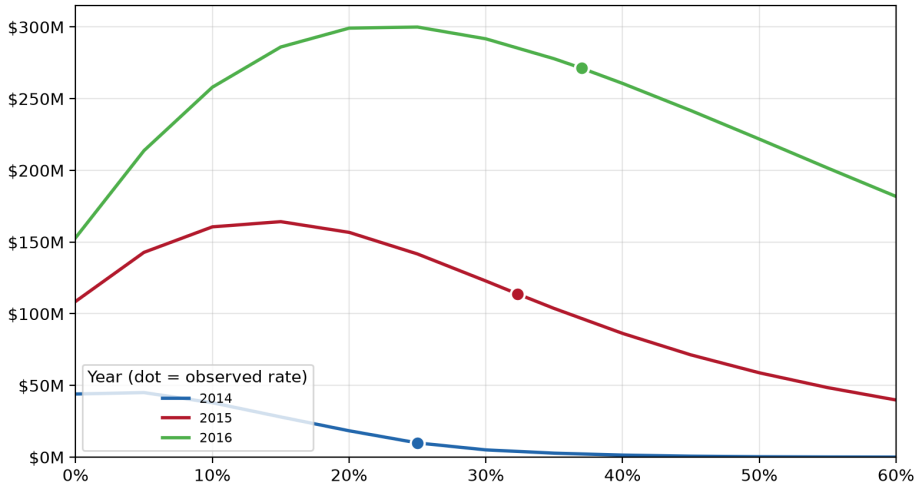


Figure 20: Nested logit, Model 6: counterfactual tax revenue by excise rate with illegal purchases placed in the outside good, 2016.

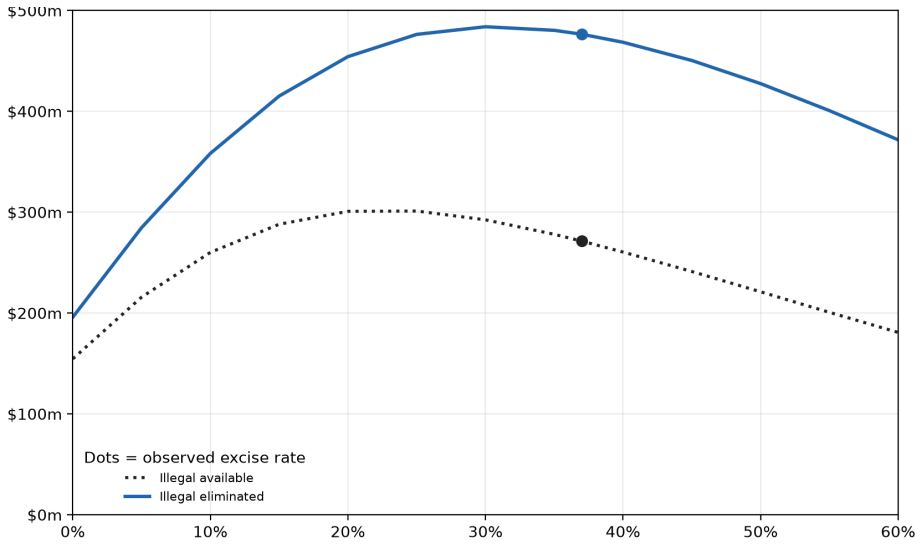


Figure 21: Nested logit, Model 6: counterfactual tax revenue by excise rate with illegal sales banned, 2016.

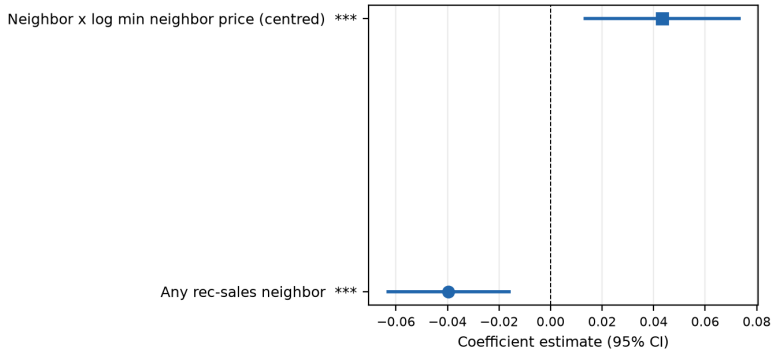


Figure 22: Effect of Neighboring-State Recreational Sales on the Log Illegal Price: Two-Way Fixed Effects Estimates of Equation (36)

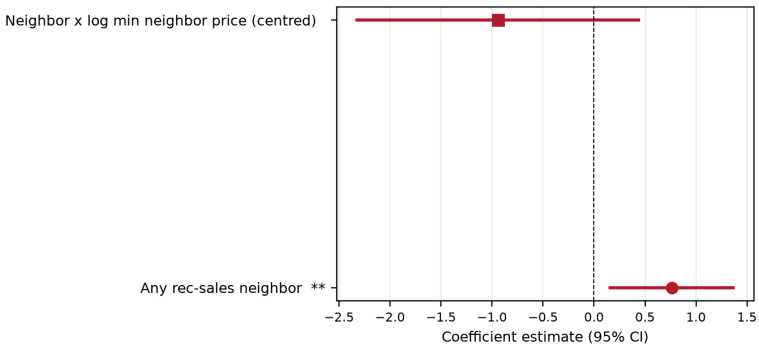


Figure 23: Effect of Neighboring-State Recreational Sales on Past-Year Use: Two-Way Fixed Effects Estimates of Equation (36)

Table 11: Event-Study Estimates: Neighbor and Own-State Recreational Sales, Illegal Prices and Cannabis Use, State Panel 2011–2024

	(1)	(2)	(3)
	Log illegal price	Past-year use	Past-month use
Panel A. Neighbor event study			
Event time ≤ -4	0.0235*	−0.3472	−0.0703
	(0.0120)	(0.3543)	(0.3544)
Event time −3	0.0125	−0.0591	0.0313
	(0.0076)	(0.2638)	(0.2335)
Event time −2	0.0046	−0.0522	0.0498
	(0.0039)	(0.1492)	(0.1267)
Event time −1		— reference —	
Event time 0	−0.0092**	0.1993	0.0584
	(0.0044)	(0.2035)	(0.2051)
Event time 1	−0.0173**	0.2391	−0.0102
	(0.0073)	(0.3162)	(0.2989)
Event time 2	−0.0177*	0.2090	−0.0287
	(0.0102)	(0.3484)	(0.3115)
Event time 3	−0.0240**	0.7936*	0.3040
	(0.0121)	(0.4471)	(0.3908)
Event time ≥ 4	−0.0542***	1.2694**	0.5197
	(0.0165)	(0.4997)	(0.4671)
Observations / states	663 / 51	714 / 51	714 / 51
Model R^2	0.171	0.320	0.294
Pre-trend p-value	0.259	0.715	0.919
Panel B. Own-state event study			
Event time ≤ -4	0.0201	−1.7330***	−1.4776***
	(0.0149)	(0.4700)	(0.4419)
Event time −3	0.0151	−1.0007***	−0.9152***
	(0.0142)	(0.2423)	(0.2656)
Event time −2	0.0076	−0.7613***	−0.7072***
	(0.0078)	(0.1979)	(0.1922)
Event time −1		— reference —	
Event time 0	−0.0070	0.8192***	0.7732***
	(0.0043)	(0.2712)	(0.2349)
Event time 1	−0.0073	1.7299***	1.6689***
	(0.0079)	(0.4740)	(0.4528)
Event time 2	−0.0108	2.2959***	1.8245***
	(0.0119)	(0.6475)	(0.5329)
Event time 3	−0.0287*	2.9237***	2.2087***
	(0.0151)	(0.7073)	(0.5466)
Event time ≥ 4	−0.0512***	3.4714***	2.6352***
	(0.0142)	(0.7927)	(0.6249)
Observations / states	663 / 51	714 / 51	714 / 51
Model R^2	0.107	0.309	0.275
Pre-trend p-value	0.608	< 0.001	0.001

Notes: Each column reports estimates of Equation (38) from a separate regression. The dependent variables are the log illegal retail price (column 1) and National Survey on Drug Use and Health (NSDUH) past-year (column 2) and past-month (column 3) prevalence of cannabis use. Standard errors clustered at the state level are in parentheses. All specifications include a constant, state and year fixed effects, and the controls $\mathbf{X}_{st}$. Event time −1 is the omitted reference period, and event times ≤ -4 and ≥ 4 are binned; the pre-trend p-value is from a joint test of the pre-treatment coefficients. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 12: Tax-Model Estimates: Recreational Sales, Excise Taxes, Illegal Prices and Cannabis Use, State Panel 2011–2024

	(1)	(2)	(3)
	Log illegal price	Past-year use	Past-month use
Panel A. Neighbor tax model			
Any recreational-sales neighbor	−0.0562***	0.7353	0.5754
	(0.0179)	(0.7144)	(0.5957)
Neighbor × ln(1 + population-weighted neighbor tax)	0.1799**	−1.4964	−2.7760
	(0.0891)	(3.9617)	(3.3954)
Observations / states	663 / 51	714 / 51	714 / 51
Model R^2	0.126	0.299	0.290
Panel B. Own-state tax model			
Recreational sales active	−0.0064	3.0810**	2.7256**
	(0.0217)	(1.2211)	(1.0576)
Recreational sales × ln(1 + τ_{st})	−0.1238	−0.3144	−0.9081
	(0.0918)	(5.5641)	(4.7817)
Observations / states	663 / 51	714 / 51	714 / 51
Model R^2	0.068	0.235	0.221

Notes: Each column reports estimates from a separate regression; Panel A estimates Equation (40) and Panel B Equation (39). The dependent variables are the log illegal retail price (column 1) and National Survey on Drug Use and Health (NSDUH) past-year (column 2) and past-month (column 3) prevalence of cannabis use. Standard errors clustered at the state level are in parentheses. All specifications include a constant, state and year fixed effects, and the controls $\mathbf{X}_{st}$. The tax interactions are not centered. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

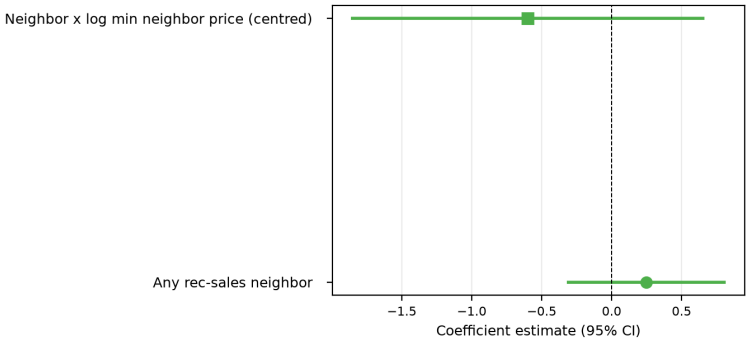


Figure 24: Effect of Neighboring-State Recreational Sales on Past-Month Use: Two-Way Fixed Effects Estimates of Equation (36)

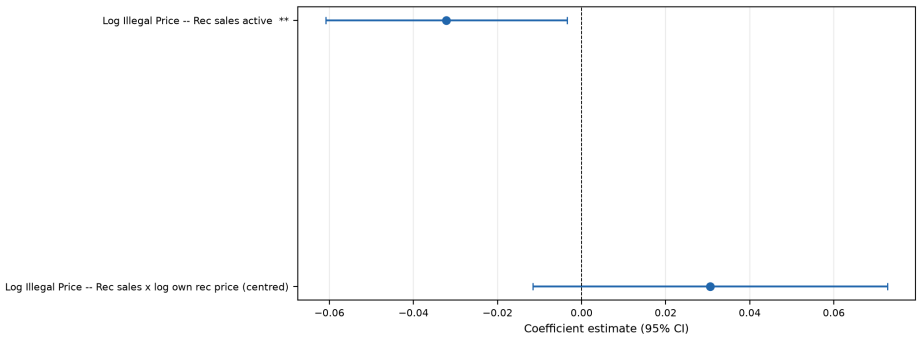


Figure 25: Effect of Own-State Recreational Sales on the Log Illegal Price: Two-Way Fixed Effects Estimates of Equation (37)

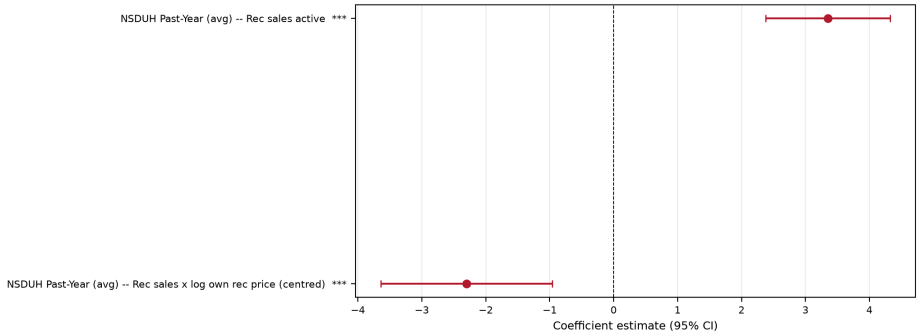


Figure 26: Effect of Own-State Recreational Sales on Past-Year Use: Two-Way Fixed Effects Estimates of Equation (37)

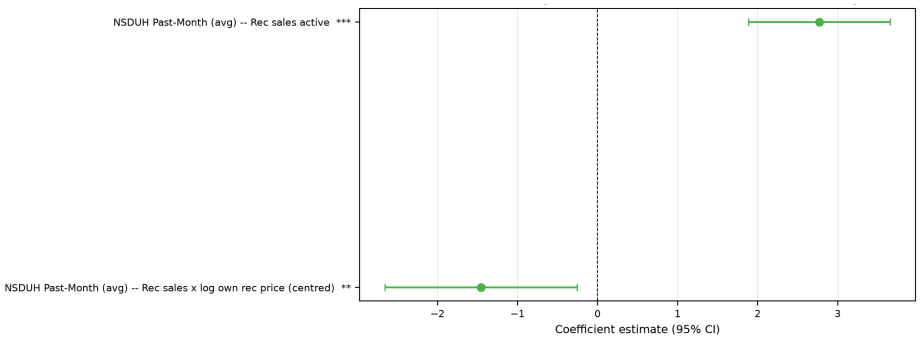


Figure 27: Effect of Own-State Recreational Sales on Past-Month Use: Two-Way Fixed Effects Estimates of Equation (37)

A.0.2 Reduced Form Estimates and Figures

A.0.3 Maps, Regulatory Timeline and Summary Statistics

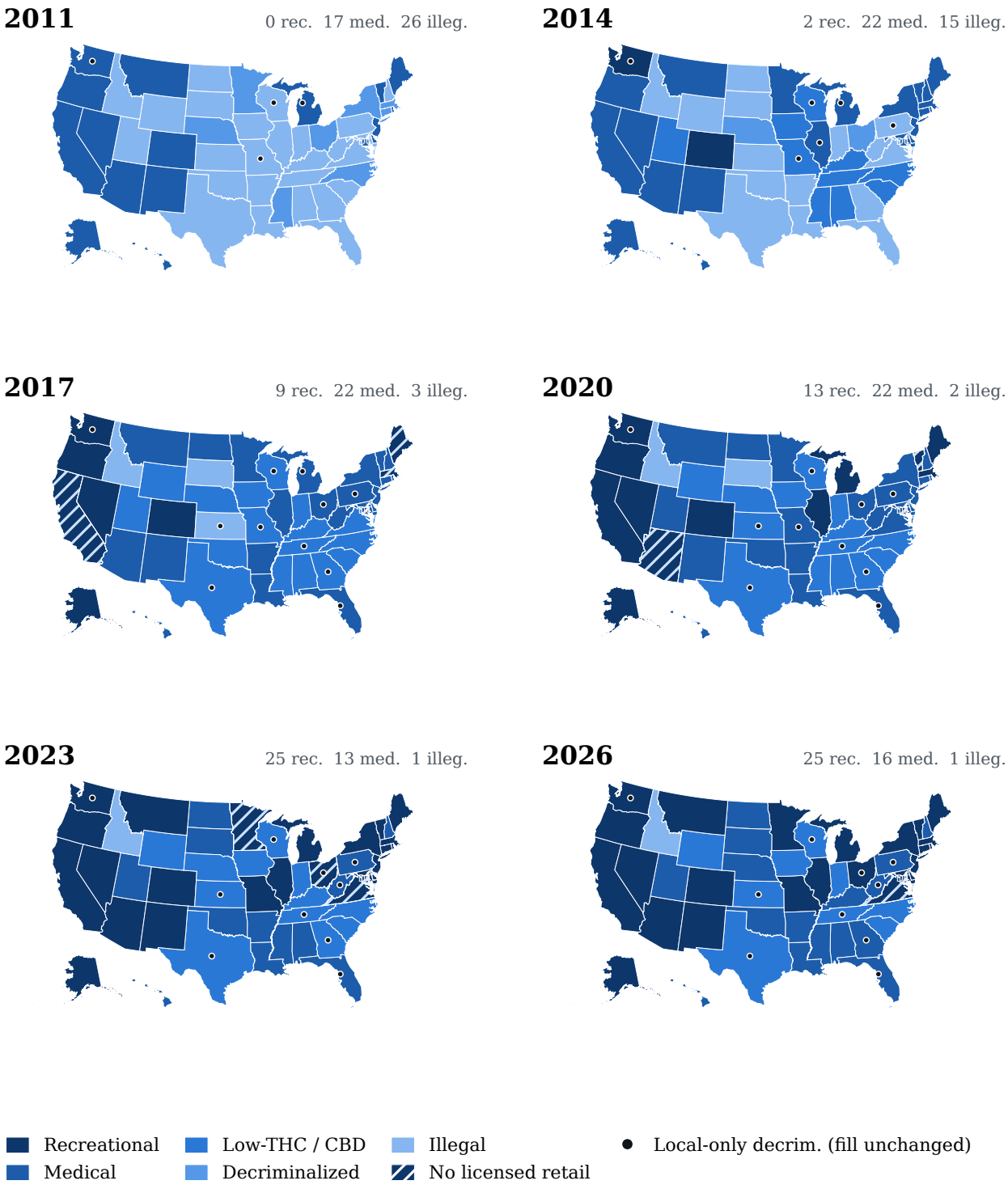


Figure 28: Cannabis legal status by state in six snapshot years between 2011 and 2024, read left to right and top to bottom. Fill colors give each state's legal status; hatching marks legal possession without a licensed retail market; dots mark local-only decriminalization, which does not change a state's fill.

Table 13: Washington State cannabis regulatory actions, 1923–2024.

Date	Type	Action
Mar 1923	Prohibition	**Washington Narcotics Act of 1923: cannabis defined as a narcotic drug** Washington's 1923 Narcotics Act added 'cannabis americana' and 'cannabis indica' to the state's list of prohibited narcotic drugs.
Nov 1998	Medical	**Initiative 692: Washington State Medical Use of Marijuana Act** Voters approved an initiative giving qualifying patients and their designated primary caregivers an affirmative defense, not an affirmative right, to state marijuana charges. *Effective 3 Dec 1998.*
Oct 2008	Medical	**DOH rule defines the 60-day supply: 24 ounces and 15 plants** The Department of Health (DOH) defined the 1998 initiative's 'sixty-day supply' with a presumptive limit of 24 ounces of usable marijuana and 15 plants. *Effective 2 Nov 2008.*
Apr 2011	Medical	**E2SSB 5073 signed with a partial veto by Gov. Christine Gregoire** Gregoire signed the medical cannabis overhaul on April 29, 2011, but vetoed 36 sections, removing the state-licensed dispensary, producer and processor system and the DOH patient registry, after the state's two U.S. Attorneys warned that state employees administering it could face federal prosecution. *Effective 22 Jul 2011.*
Jul 2011	Enforcement	**Collective-garden gray market emerges after the E2SSB 5073 veto** With licensed dispensaries vetoed but collective gardens legal, hundreds of storefront 'access points' opened across Washington operating as rotating collective gardens, an unregulated, untaxed medical market alongside I-502.
Nov 2012	Legalization	**Initiative 502: legalization and regulation of adult-use marijuana** Voters legalized possession of limited amounts of marijuana by adults 21+, directed the Liquor Control Board (LCB) to license a three-tier producer/processor/retailer industry, and set a 5 ng/mL THC per se DUI limit. *Effective 6 Dec 2012.*
Dec 2012	Tax	**I-502's original three-tier 25% excise tax** I-502 imposed a 25% excise tax at each of three transfer points (producer to processor, processor to retailer, and retailer to consumer), which compounded into an effective rate far above 25%.
Jul 2014	Legalization	**First legal recreational cannabis sales in Washington** Washington's first licensed recreational retailers opened on July 8, 2014, roughly 20 months after I-502 passed.
Apr 2015	Medical	**SB 5052, Cannabis Patient Protection Act: medical market folded into the regulated system** SB 5052 merged Washington's unregulated medical cannabis market into the I-502 licensing system, ending collective gardens as of July 1, 2016 and creating a voluntary patient database, medical endorsements for retailers, and recognition cards. *Effective 24 Jul 2015.*
Jun 2015	Tax	**HB 2136: excise tax restructured to a single 37% retail tax** HB 2136 replaced I-502's stacked three-tier 25% excise tax with one 37% excise tax collected at retail, and added local revenue sharing plus licensing changes. *Effective 1 Jul 2015.*
2016	Licensing	**Retail license cap raised from 334 to 556 stores** To absorb medical patient demand after the merger, the LCB increased the statewide retail cap and opened a second application window with priority for former medical operators.

continued on next page

Table 13, continued

Date	Type	Action
Jul 2016	Medical	**Collective gardens end; four-patient cooperatives begin** Washington's unregulated collective gardens and the storefront dispensaries built on them became illegal on July 1, 2016, replaced by tightly limited four-patient cooperatives and medically endorsed retail stores.
Mar 2024	Tax	**HB 1453: 37% cannabis excise tax exemption for registered medical patients** Washington exempted qualifying medical patients with a DOH recognition card from the 37% cannabis excise tax on DOH-compliant products bought at medically endorsed retailers through June 30, 2029. *Effective 6 Jun 2024.*

Table 14: Summary Statistics for the Washington Estimation Sample, 2014–2016

Variable	2014	2015	2016
Instrument source variables			
Available brands^{L}	8.462 (3.912)	21.348 (9.469)	32.553 (13.861)
Flower share^{L}	0.896 (0.083)	0.885 (0.051)	0.835 (0.054)
Weekly wholesale price^{L}	7.618 (1.338)	4.562 (0.615)	4.159 (0.492)
County average temperature^{M}	59.421 (3.254)	60.006 (3.235)	59.124 (3.071)
Market size M_{st} ^{M}	6.496×10^6 (1.316×10^7)	6.590×10^6 (1.341×10^7)	6.722×10^6 (1.374×10^7)
I-502 vote share^{M}	0.506 (0.068)	0.506 (0.068)	0.506 (0.068)
Illegal observation (share)^{R}	0.328 (0.471)	0.168 (0.375)	0.104 (0.306)
Entry-cost shifters			
Police officers^{M}	270.026 (584.034)	275.333 (598.514)	279.615 (610.209)
Lagged property incidents^{M}	45,253.897 (24,785.299)	43,713.795 (24,155.180)	42,929.513 (23,908.436)
Arrest indicators			
Sales and manufacturing arrests^{M}	470.872 (261.620)	407.410 (278.754)	406.744 (228.752)
Possession arrests^{M}	3,410.051 (1,905.058)	3,311.769 (1,700.946)	3,426.385 (1,700.902)
Controls and fixed effects			
Local sales tax rate^{L}	0.022 (0.006)	0.024 (0.006)	0.025 (0.007)
Excise tax rate^{L}	0.250 (0.000)	0.323 (0.000)	0.370 (0.000)
State sales tax rate^{L}	0.065 (0.000)	0.065 (0.000)	0.065 (0.000)
Region fixed-effect categories^{M}	8	8	8
Year fixed effects	Included	Included	Included

Notes: Values are annual means with sample standard deviations in parentheses. Superscripts give the sample: R, all legal products plus one illegal observation per county-year; L, legal products only; M, one observation per county-year market. The annual sample sizes are $N_R = (119, 232, 375)$, $N_L = (80, 193, 336)$, and $N_M = (39, 39, 39)$. The 2014 excise rate is the retail-stage statutory rate.

Table 15: Summary Statistics for the State Cannabis Policy Panel, 2011–2017

Variable	2011	2012	2013	2014	2015	2016	2017
Outcomes							
Log Illegal Cannabis Price	–	2.381 (0.094)	2.338 (0.094)	2.315 (0.082)	2.262 (0.076)	2.237 (0.072)	2.186 (0.081)
Log Recreational Cannabis Price	–	–	–	2.686 (0.251)	2.393 (0.102)	2.341 (0.306)	2.262 (0.348)
Log Medical Cannabis Price	–	–	–	2.673 (0.309)	2.754 (0.461)	2.678 (0.453)	2.655 (0.494)
Past-Year Cannabis Use, 18+ (Avg., p.p.)	11.538 (2.945)	12.021 (3.208)	12.637 (3.451)	13.327 (3.450)	13.796 (3.764)	14.258 (4.204)	15.396 (4.589)
Past-Month Cannabis Use, 18+ (Avg., p.p.)	7.054 (2.220)	7.257 (2.364)	7.614 (2.551)	8.271 (2.630)	8.733 (2.842)	9.199 (3.328)	10.115 (3.749)
Past-Year Cannabis Use, 18+ (Raw, p.p.)	11.538 (2.945)	12.021 (3.208)	12.637 (3.451)	13.327 (3.450)	13.796 (3.764)	14.258 (4.204)	15.396 (4.589)
Past-Month Cannabis Use, 18+ (Raw, p.p.)	7.054 (2.220)	7.257 (2.364)	7.614 (2.551)	8.271 (2.630)	8.733 (2.842)	9.199 (3.328)	10.115 (3.749)
Cannabis Consumption (Grams per Capita)	19.106 (5.898)	19.679 (6.286)	20.652 (6.785)	22.372 (6.981)	23.579 (7.543)	24.951 (9.227)	27.545 (10.775)
Controls							
Medical Legalization Indicator	0.333 (0.476)	0.353 (0.483)	0.392 (0.493)	0.471 (0.504)	0.490 (0.505)	0.569 (0.500)	0.588 (0.497)
Active Recreational Sales Indicator	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.039 (0.196)	0.059 (0.238)	0.078 (0.272)	0.098 (0.300)
Log(1 + Rec. Dispensaries per 100,000)‡	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.040 (0.246)	0.104 (0.435)	0.145 (0.530)	0.193 (0.614)
Log Population	15.124 (1.042)	15.132 (1.041)	15.139 (1.040)	15.145 (1.040)	15.151 (1.040)	15.157 (1.041)	15.163 (1.042)
Unemployment Rate (p.p.)	8.163 (1.932)	7.369 (1.716)	6.765 (1.539)	5.782 (1.282)	5.043 (1.079)	4.676 (0.995)	4.176 (0.925)
Poverty Rate (p.p.)	14.308 (3.192)	14.304 (3.420)	14.010 (3.554)	14.063 (3.907)	12.939 (3.054)	12.453 (3.113)	12.129 (2.911)
Log Gross State Product per Capita	-3.017 (0.265)	-2.988 (0.266)	-2.967 (0.259)	-2.934 (0.260)	-2.909 (0.257)	-2.897 (0.256)	-2.865 (0.254)
Food Insecurity Rate (p.p.)	15.665 (3.051)	15.562 (3.096)	15.296 (3.580)	15.009 (3.232)	13.522 (3.140)	13.186 (3.279)	12.320 (2.543)
Very Low Food Security Rate (p.p.)	5.389 (1.261)	5.560 (1.258)	5.229 (1.480)	5.529 (1.406)	4.694 (1.482)	4.774 (1.623)	4.075 (1.211)
Log(1 + DEA Plants per 100,000)	5.909 (1.840)	5.206 (2.228)	5.245 (1.699)	5.366 (1.651)	4.872 (2.074)	4.543 (2.242)	4.501 (2.019)
Log(1 + DEA Outdoor Sites per 100,000)	1.153 (0.963)	0.866 (0.804)	0.771 (0.774)	0.845 (0.760)	0.760 (0.749)	0.679 (0.710)	0.622 (0.686)
Log(1 + DEA Bulk Pounds per 100,000)	1.766 (1.568)	1.744 (1.472)	1.930 (1.636)	2.067 (1.504)	1.908 (1.437)	2.025 (1.561)	2.223 (1.503)
Log(1 + DEA Assets per Capita)	0.096 (0.123)	0.078 (0.114)	0.066 (0.098)	0.076 (0.108)	0.071 (0.099)	0.090 (0.159)	0.061 (0.111)
DEA Indoor Share of Plants	0.330 (0.318)	0.233 (0.272)	0.292 (0.307)	0.278 (0.302)	0.247 (0.302)	0.292 (0.312)	0.332 (0.339)
Log(1 + DEA Indoor Sites per 100,000)	0.701 (0.544)	0.488 (0.531)	0.514 (0.475)	0.462 (0.436)	0.336 (0.339)	0.319 (0.338)	0.295 (0.348)
Log(1 + DEA Weapons per 100,000)	0.832 (0.663)	0.801 (0.964)	0.642 (0.588)	0.577 (0.644)	0.516 (0.531)	0.502 (0.492)	0.545 (0.540)

Notes: Cells report annual means with sample standard deviations in parentheses. The balanced panel covers the 50 states and the District of Columbia, observed annually from 2011 through 2024. Use is in percentage points among adults 18 and older; the raw NSDUH series excludes 2020–2021, and the averaged series supplies values for those years. The DEA measures are available for 574 of the 714 state-years. ‡ Included in the neighbor designs only, since it may be a post-treatment outcome in the own-state designs.

Table 16: Summary Statistics for the State Cannabis Policy Panel, 2018–2024

Variable	2018	2019	2020	2021	2022	2023	2024
Outcomes							
Log Illegal Cannabis Price	2.154 (0.081)	2.107 (0.087)	2.061 (0.092)	2.023 (0.104)	1.970 (0.116)	1.926 (0.112)	1.869 (0.120)
Log Recreational Cannabis Price	2.200 (0.375)	2.127 (0.378)	2.256 (0.503)	2.169 (0.452)	2.233 (0.430)	2.180 (0.399)	2.119 (0.391)
Log Medical Cannabis Price	2.625 (0.492)	2.644 (0.564)	2.597 (0.545)	2.540 (0.533)	2.403 (0.516)	2.322 (0.523)	2.252 (0.522)
Past-Year Cannabis Use, 18+ (Avg., p.p.)	16.556 (4.763)	17.663 (4.836)	18.790 (4.827)	21.188 (5.021)	22.149 (5.303)	23.752 (5.375)	24.453 (5.615)
Past-Month Cannabis Use, 18+ (Avg., p.p.)	10.845 (3.750)	11.539 (3.733)	12.644 (3.812)	14.650 (4.037)	15.546 (4.319)	16.865 (4.456)	17.271 (4.421)
Past-Year Cannabis Use, 18+ (Raw, p.p.)	16.556 (4.763)	17.663 (4.836)	–	–	22.149 (5.303)	23.752 (5.375)	24.453 (5.615)
Past-Month Cannabis Use, 18+ (Raw, p.p.)	10.845 (3.750)	11.539 (3.733)	–	–	15.546 (4.319)	16.865 (4.456)	17.271 (4.421)
Cannabis Consumption (Grams per Capita)	30.247 (11.152)	32.257 (11.366)	35.510 (12.677)	41.813 (13.010)	44.008 (12.732)	47.441 (12.600)	48.432 (12.280)
Controls							
Medical Legalization Indicator	0.667 (0.476)	0.667 (0.476)	0.706 (0.460)	0.725 (0.451)	0.745 (0.440)	0.765 (0.428)	0.765 (0.428)
Active Recreational Sales Indicator	0.137 (0.348)	0.157 (0.367)	0.196 (0.401)	0.216 (0.415)	0.333 (0.476)	0.392 (0.493)	0.412 (0.497)
Log(1 + Rec. Dispensaries per 100,000)‡	0.232 (0.680)	0.258 (0.711)	0.338 (0.775)	0.406 (0.852)	0.607 (1.055)	0.723 (1.105)	0.772 (1.129)
Log Population	15.168 (1.043)	15.172 (1.044)	15.183 (1.044)	15.186 (1.043)	15.191 (1.043)	15.199 (1.043)	15.207 (1.044)
Unemployment Rate (p.p.)	3.794 (0.831)	3.622 (0.822)	7.367 (1.844)	4.835 (1.260)	3.394 (0.756)	3.318 (0.762)	3.712 (0.817)
Poverty Rate (p.p.)	11.459 (3.045)	10.233 (2.985)	11.235 (2.766)	11.175 (3.173)	11.090 (3.195)	10.555 (2.946)	10.200 (2.928)
Log Gross State Product per Capita	-2.819 (0.253)	-2.787 (0.254)	-2.799 (0.254)	-2.698 (0.249)	-2.604 (0.243)	-2.546 (0.239)	-2.506 (0.237)
Food Insecurity Rate (p.p.)	11.627 (2.980)	11.053 (2.577)	11.157 (2.708)	10.266 (2.466)	13.033 (3.239)	13.272 (3.179)	13.965 (3.441)
Very Low Food Security Rate (p.p.)	4.100 (1.209)	3.824 (1.107)	3.805 (1.261)	3.253 (1.090)	4.712 (1.656)	4.945 (1.425)	4.980 (1.582)
Log(1 + DEA Plants per 100,000)	3.905 (2.344)	3.464 (2.551)	3.987 (2.278)	3.899 (2.463)	3.115 (2.705)	3.315 (2.988)	3.634 (2.915)
Log(1 + DEA Outdoor Sites per 100,000)	0.511 (0.636)	0.534 (0.625)	0.550 (0.663)	0.522 (0.660)	0.454 (0.663)	0.287 (0.888)	0.817 (1.284)
Log(1 + DEA Bulk Pounds per 100,000)	3.035 (2.078)	3.165 (1.905)	2.503 (2.074)	2.961 (2.114)	2.788 (2.034)	3.155 (2.235)	2.992 (2.388)
Log(1 + DEA Assets per Capita)	0.113 (0.236)	0.070 (0.095)	0.108 (0.149)	0.194 (0.327)	0.117 (0.197)	0.110 (0.211)	0.147 (0.299)
DEA Indoor Share of Plants	0.371 (0.367)	0.326 (0.356)	0.343 (0.372)	0.231 (0.325)	0.347 (0.402)	0.394 (0.417)	0.278 (0.390)
Log(1 + DEA Indoor Sites per 100,000)	0.302 (0.369)	0.275 (0.320)	0.236 (0.300)	0.243 (0.307)	0.197 (0.285)	0.177 (0.266)	0.330 (0.482)
Log(1 + DEA Weapons per 100,000)	0.596 (0.600)	0.503 (0.495)	0.464 (0.476)	0.519 (0.671)	0.439 (0.533)	0.594 (0.731)	0.642 (0.661)

Notes: See the notes to Table 15.

A.1 Aggregative Games: Proofs and Lemmas

Proof of Proposition 1S. Fix a prefix P and a mass N. Legal firms play $\widetilde{r}_\ell(A)$ and illegal sellers $\widehat{r}_f$ (Lemma 1), so the aggregate solves

$$\frac{1 + N\widehat{r}_f}{A} + \sum_{\ell \in P} \widetilde{s}_\ell(A) = 1.$$

The left side falls strictly from a value of at least one at $A = 1 + N\widehat{r}_f$ to zero, so the root is unique. An illegal seller earns $\mu\widehat{r}_f/A$, which free entry sets to K_F when $N > 0$; this gives (9) (Lemma 3(i)). Every prefix therefore has one continuation, and A5(b) has nothing to select. Within a prefix all legal firms face the same A^*, and $\bar{\pi}_\ell(A) = \mu\widetilde{s}_\ell/(1 - \widetilde{s}_\ell)$ rises with the margin and falls with A; since adding a firm weakly raises A^*, A5(c) gives (D1) and (D2) (Lemma 3(ii)), and Lemma 2 selects P_{L^*}, with $\Gamma_L > 1$ exactly for $L \le L^*$. Write $\widetilde{A} = \widetilde{A}(P_{L^*})$, with $\widetilde{A}(P_0) = 1$; a tie $\widetilde{A} = \bar{A}$ goes to $N^* = 0$.

Case 1: Legal firms and illegal sellers enter ($L^* > 0$, $\widetilde{A} < \bar{A}$):

$$A^* = \bar{A}, \qquad a_\ell = \widetilde{r}_\ell(\bar{A}), \qquad N^* = \frac{\bar{A} - 1 - \sum_{\ell \in P_{L^*}} a_\ell}{\widehat{r}_f} > 0. \tag{41}$$

Free entry binds and pins the aggregate at $\bar{A}$, which contains only illegal primitives, so consumer surplus is $\mu \ln \bar{A}$. Each legal firm prices with N already fixed and plays its inclusive reply at $\bar{A}$ (Lemma 1(ii)); the fringe fills the rest, $A_F = \bar{A} - 1 - A_L$. Legal primitives and the tax therefore set only the composition of the aggregate.

Case 2: Legal firms enter and illegal sellers do not ($L^* > 0$, $\widetilde{A} \ge \bar{A}$):

$$A^* = \widetilde{A}, \qquad a_\ell = \widetilde{r}_\ell(\widetilde{A}), \qquad N^* = 0. \tag{42}$$

Legal firms alone reach $\bar{A}$, so any positive mass would push A above $\widetilde{A} \ge \bar{A}$ and each seller would earn less than K_F. The aggregate is the legal-only root of (7) with $N^* = 0$ and is set by legal primitives. Because prices are chosen after N is fixed, each legal payoff is smooth in its own action, and there is no limit-pricing subcase.

Case 3: Illegal sellers enter and legal firms do not ($L^* = 0$, $\bar{A} > 1$):

$$A^* = \bar{A}, \qquad N^* = \frac{\bar{A} - 1}{\widehat{r}_f} > 0. \tag{43}$$

No candidate covers its fixed cost: $\Gamma_L \le 1$ for every L, which under (D2) reduces to $\Gamma_1 \le 1$. Against the outside good alone the first seller earns $\mu\widehat{r}_f$, which exceeds K_F exactly when $\bar{A} > 1$, so sellers

enter until $A = \bar{A}$. This case uses only Lemma 1(i) and the zero-profit condition.

Case 4: Neither enters ($L^* = 0$, $\bar{A} \leq 1$): $A^* = 1$ and $N^* = 0$. A seller would earn $\mu\widehat{r}_f \leq K_F$ even against the outside good alone, and, as in Case 3, $\Gamma_L \leq 1$ for every L. Every consumer buys the outside good, so $s_0 = 1$ and consumer surplus is zero.

The cases are exhaustive and mutually exclusive: given L^*, the fringe is active exactly when $\widetilde{A} < \bar{A}$, which for $L^* = 0$ reads $\bar{A} > 1$. Whenever illegal sellers are active they charge $c_F + \mu$ and legal prices are $v_\ell - \mu \ln a_\ell$. □

Proof of Proposition 2. Fix a nonempty legal set and a range of taxes at which illegal sellers are active. By Lemma 4, a higher τ leaves the aggregate at $\bar{A}$ and consumer welfare $\mu \ln \bar{A}$ unchanged but lowers A_L and raises A_F and N. Total welfare changes through legal profits, tax revenue and the harms $C_L(A_L)$, $C_F(A_F)$ and $C_F^E(N)$, so it depends on the composition of the aggregate as well as its size. □

The naive policymaker. On the same range, the first-order condition of a policymaker who maximizes total welfare is

$$\frac{\partial \mathrm{TW}}{\partial \tau} = \frac{\partial \mathrm{CS}}{\partial \tau} + \frac{\partial \mathrm{PS}}{\partial \tau} + \frac{\partial \mathrm{TR}}{\partial \tau} - \frac{\partial C_L}{\partial A_L}\frac{\partial A_L}{\partial \tau} - \frac{\partial C_F}{\partial A_F}\frac{\partial A_F}{\partial \tau} - \frac{\partial C_F^E}{\partial N}\frac{\partial N}{\partial \tau} = 0,$$

while the naive policymaker sets the tax according to

$$\frac{\partial \mathrm{NW}}{\partial \tau} = \frac{\partial \mathrm{CS}}{\partial \tau} + \frac{\partial \mathrm{PS}}{\partial \tau} + \frac{\partial \mathrm{TR}}{\partial \tau} - \frac{\partial C_L}{\partial A_L}\frac{\partial A_L}{\partial \tau} = 0.$$

Let τ^* maximize total welfare TW and τ^n maximize naive welfare NW. Then $\mathrm{NW} - \mathrm{TW} = C_F(A_F) + C_F^E(N)$, which rises strictly with τ by Lemma 4(ii) when the harms are strictly increasing. Optimality gives $\mathrm{TW}(\tau^*) \geq \mathrm{TW}(\tau^n)$ and $\mathrm{NW}(\tau^n) \geq \mathrm{NW}(\tau^*)$; adding, $(\mathrm{NW} - \mathrm{TW})(\tau^n) \geq (\mathrm{NW} - \mathrm{TW})(\tau^*)$, so $\tau^n \geq \tau^*$. If τ^* is interior, $\mathrm{NW}'(\tau^*) = (\mathrm{NW} - \mathrm{TW})'(\tau^*) > 0$, so $\tau^n > \tau^*$: the naive policymaker over-taxes. At τ^n the legal sector is smaller and illegal entry larger than at τ^*, while the aggregate is the same (Lemma 4). The argument uses only that the illegal harms $C_F(A_F) + C_F^E(N)$ rise with τ, which Corollary 2(iv) also gives under nested logit, so for a fixed nonempty legal set the naive policymaker over-taxes in the estimated model as well.

Lemma 1 (Logit replies)**.** *Let $\widehat{r}_j \equiv e^{(m_j - \mu)/\mu}$, with $m_F \equiv v_F - c_F$, and $g(s) \equiv \ln s + \frac{s}{1-s}$ on $(0, 1)$.*

(i) An illegal seller sets $p_f = c_F + \mu$, i.e., $a_f = \widehat{r}_f$, and earns operating profit $\mu\,\widehat{r}_f/A$.

(ii) A legal firm's profit is strictly quasi-concave in its own price. Its best reply, when the aggregate is A, is the share $\widetilde{s}_\ell(A) = g^{-1}\big(\ln \widehat{r}_\ell - \ln A\big)$; equivalently $p_\ell = c_\ell + \tau + \mu/(1 - s_\ell)$ and $\widetilde{r}_\ell(A) = A\,\widetilde{s}_\ell(A)$.

(iii) $\widetilde{s}_\ell(A)$ *is strictly decreasing in* A*, strictly increasing in* m_ℓ*, and tends to zero as* $A \to \infty$*;* $0 < \widetilde{r}_\ell(A) < \widehat{r}_\ell$.

(iv) Equilibrium operating profit $\bar{\pi}_\ell(A) \equiv \mu\,\widetilde{s}_\ell(A)/(1-\widetilde{s}_\ell(A))$ *is strictly decreasing in* A *and strictly increasing in* m_ℓ.

Proof. Since $\mu \ln a_j = v_j - p_j$, firm j's markup is $m_j - \mu \ln a_j$, so p_j and a_j are inversely related. (i) At fixed A, $\partial\pi_f/\partial a_f = (m_F - \mu - \mu \ln a_f)/A$ is strictly decreasing and vanishes only at $\widehat{r}_f$, where the markup is μ. (ii) Fix the rivals, so $A_{-\ell} = A - a_\ell \geq 1$ is fixed. The left side of (6), $\big[(m_\ell - \mu \ln a_\ell)(1 - a_\ell/A) - \mu\big]/A$, has the sign of $\ln \widehat{r}_\ell - \ln a_\ell - a_\ell/A_{-\ell}$, which falls strictly from $+\infty$ to $-\infty$; profit is therefore single-peaked in a_ℓ, hence strictly quasi-concave in p_ℓ. With $s = a_\ell/A$ and $\ln a_\ell = \ln s + \ln A$, the first-order condition $(m_\ell - \mu \ln a_\ell)(1-s) = \mu$ (markup $\mu/(1-s)$) reads $g(s) = \ln \widehat{r}_\ell - \ln A$, and g is a strictly increasing bijection of $(0,1)$ onto $\mathbb{R}$ ($g'(s) = 1/s + 1/(1-s)^2$), so the reply has share $\widetilde{s}_\ell(A)$.[14] (iii) The argument $(m_\ell - \mu)/\mu - \ln A$ of g^{-1} falls in A, to $-\infty$, and rises in m_ℓ, while g^{-1} is increasing and tends to zero at $-\infty$. Adding $\ln A$ to the first-order condition gives $\ln \widetilde{r}_\ell(A) = \ln \widehat{r}_\ell - \widetilde{s}_\ell/(1-\widetilde{s}_\ell) < \ln \widehat{r}_\ell$. (iv) Profit is markup times share, $\mu\widetilde{s}_\ell/(1-\widetilde{s}_\ell)$, increasing in $\widetilde{s}_\ell$; apply (iii). Finally, differentiating $g(\widetilde{s}_\ell(A)) = \ln \widehat{r}_\ell - \ln A$ gives, with $s = \widetilde{s}_\ell(A)$,

$$\frac{A\,\widetilde{r}'_\ell(A)}{\widetilde{r}_\ell(A)} = 1 - \frac{1}{s\,g'(s)} = \frac{s}{1 - s + s^2} \in (0,1),$$

which is A3–A4; A2 is (i)–(ii), and A1 holds because $\partial\pi_j/\partial A = -(\text{markup})\,a_j/A^2 < 0$ at the replies. □

Lemma 2 (Rank-ordered entry)**.** *Suppose every legal prefix* P_L*,* $L = 0, \ldots, \bar{L}$*, reached or not, has a unique (or uniquely selected) continuation. Then rank-ordered entry admits a unique legal set, and it is a prefix* P_{L^*}*. If, in addition,*

$$\textit{(D1)} \qquad \Gamma_k(P_L) \geq \Gamma_L(P_L) \qquad \textit{for all } k < L \leq \bar{L},$$

then $L^* = \max\big(\{0\} \cup \{L : \Gamma_L > 1\}\big)$*, so* $\Gamma_L \leq 1$ *for every* $L > L^*$ *and every entrant's ratio in the final set exceeds one. If also*

$$\textit{(D2)} \qquad \Gamma_{L+1} < \Gamma_L \qquad \textit{for all } 1 \leq L < \bar{L},$$

then $\Gamma_L > 1$ *exactly for* $L \leq L^*$*: if* $L^* \geq 1$*,* L^* *is the unique index with* $\Gamma_{L^*} > 1 \geq \Gamma_{L^*+1}$ *(with* $\Gamma_{\bar{L}+1} \equiv 0$*), and if* $\bar{L} \geq 1$*,* $L^* = 0$ *exactly when* $\Gamma_1 \leq 1$.

[14] With $x = a_\ell/A_{-\ell}$ the first-order condition reads $xe^x = \widehat{r}_\ell/A_{-\ell}$, so the reply to rivals' total $A_{-\ell}$ is $a_\ell = A_{-\ell}\,W(\widehat{r}_\ell/A_{-\ell})$, W the Lambert W function, with markup $\mu[1 + W(\widehat{r}_\ell/A_{-\ell})]$. The inclusive reply $\widetilde{r}_\ell(A)$ has no closed form, but its inverse is explicit: $A = \widehat{r}_\ell/(s\,e^{s/(1-s)})$ with $s = \widetilde{s}_\ell(A)$.

Proof. With one continuation per prefix, each candidate's payoff from every terminal legal set is a single number, so rank-ordered entry is a finite game of perfect information. With ties broken by the stay-out convention, backward induction gives a unique outcome, a prefix because a decline ends the queue. Under (D1), induct downward from $\bar{L}$. A candidate $L > L^*$ that enters is followed only by declines, so its ratio is $\Gamma_L \leq 1$ and it declines. For $L^* \geq 1$, candidate L^* ends in P_{L^*} if it enters, with ratio $\Gamma_{L^*} > 1$, so it enters; and, by the same downward induction, each $k < L^*$ ends in P_{L^*} if it enters, with ratio $\Gamma_k(P_{L^*}) \geq \Gamma_{L^*} > 1$, so it enters. Under (D2), $\Gamma_L > 1$ exactly for $L \leq L^*$. □

Remark 1. Uniqueness in Lemma 2 uses neither (D1) nor (D2), and the form of the legal set uses only (D1), not a falling marginal ratio. Rank-ordered entry therefore pins down the legal set even under B6 (the minimum-mass assumption of Appendix A.2), where Γ^N can rise because legal firms jointly exclude the fringe. In each game below, A5(c), or B5(c) under nested logit, delivers (D1); for the deterrence profiles of the nested leader game with B6, this is Lemma 8.

Lemma 3 (Continuation and legal ranking)**.** *Fix any set P of legal candidates, possibly empty. Under logit demand:*

(i) For every $N \geq 0$ the pricing stage has a unique equilibrium: legal firms play $\widetilde{r}_\ell(A)$ and illegal sellers $\widehat{r}_f$, at the unique root $A(N;P)$ of $\frac{1+N\widehat{r}_f}{A} + \sum_{\ell\in P} \widetilde{s}_\ell(A) = 1$, which is continuous, strictly increasing and unbounded in N, weakly increasing in P, and equal to $\widetilde{A}(P)$ at $N = 0$. Free entry then selects a unique mass $N^*(P) = \max\{0,\ \bar{A} - 1 - \sum_{\ell\in P} \widetilde{r}_\ell(\bar{A})\}/\widehat{r}_f$, positive if and only if $\widetilde{A}(P) < \bar{A}$, and $A^*(P) = \max\{\bar{A}, \widetilde{A}(P)\}$.

(ii) Under A5(c), (D1) and (D2) hold.

Proof. (i) With N fixed, $\partial A/\partial a_\ell = 1$, so by A2 each legal firm's unique best reply is $\widetilde{r}_\ell(A)$; each illegal seller cannot move A and plays $\widehat{r}_f$. An equilibrium aggregate solves

$$\Psi(A;N) \equiv \frac{1+N\widehat{r}_f}{A} + \sum_{\ell\in P} \frac{\widetilde{r}_\ell(A)}{A} = 1.$$

By A4, Ψ is strictly decreasing in A; it is at least one at $A = 1 + N\widehat{r}_f$ and tends to zero as $A \to \infty$, because $\widetilde{r}_\ell < \widehat{r}_\ell$ (Lemma 1(iii)). So there is exactly one root $A(N;P) \geq 1 + N\widehat{r}_f$ (hence unbounded in N), continuous and strictly increasing in N because Ψ is, and weakly increasing in P because a firm adds a positive term. An illegal seller earns $\mu\widehat{r}_f/A(N;P)$, strictly decreasing in N and equal to K_F exactly at $\bar{A}$. If $\widetilde{A}(P) < \bar{A}$, exactly one $N^* > 0$ gives $A(N^*;P) = \bar{A}$, and $N = 0$ fails because the first entrant earns $\mu\widehat{r}_f/\widetilde{A}(P) > K_F$. If $\widetilde{A}(P) \geq \bar{A}$, every $N > 0$ gives $A > \bar{A}$ and a loss, so $N^* = 0$. By A4, $\bar{A} - 1 - \sum_{\ell\in P} \widetilde{r}_\ell(\bar{A}) > 0$ if and only if $\widetilde{A}(P) < \bar{A}$, and market clearing at $\bar{A}$ gives $N^*(P)$.

(ii) Every firm in P_L faces $A^*(P_L)$ and earns $\bar{\pi}_\ell(A^*(P_L))$, which rises with m_ℓ and falls with A (Lemma 1(iv)). Under A5(c) a lower index means a higher margin and a weakly lower fixed cost,

which gives (D1). Adding firm $L+1$ raises $\widetilde{A}$, so $A^*(P_{L+1}) \geq A^*(P_L)$ and

$$\Gamma_{L+1} = \frac{\bar{\pi}_{L+1}(A^*(P_{L+1}))}{K_{L+1}} < \frac{\bar{\pi}_L(A^*(P_{L+1}))}{K_L} \leq \frac{\bar{\pi}_L(A^*(P_L))}{K_L} = \Gamma_L,$$

which is (D2). □

The outcome in Lemma 3(i) is also the unique equilibrium if illegal entry is chosen together with prices, because one seller's entry does not move A.

A.1.1 Taxes

The counterfactuals in Section 4 change the tax and hold the licensed set fixed. Illegal sellers are active in Washington throughout the sample, so the relevant case is the one in which they are active.

Lemma 4 (A legal tax with illegal sellers active)**.** *Fix a nonempty legal set and suppose illegal sellers are active. On any range of τ on which they remain active:*

(i) Size. *$A^* = \bar{A}$: the outside share $1/\bar{A}$, consumer surplus $\mu \ln \bar{A}$ and the illegal price $c_F + \mu$ do not depend on τ.*

(ii) Composition and entry. *A higher τ lowers every legal share, and every sale the legal firms lose goes to illegal sellers: $dS_F/d\tau = -dS_L/d\tau > 0$. Illegal entry is proportional to illegal sales, $N^* = (\mu/K_F)S_F$, so it rises with τ; equivalently A_L falls and A_F rises one for one.*

(iii) Pass-through. *$dp_\ell/d\tau = \kappa_\ell \equiv \left[1 + s_\ell/(1-s_\ell)^2\right]^{-1} \in (0,1)$.*

Proof. (i) While illegal sellers are active, free entry gives $A^* = \bar{A} = \mu\widehat{r}_f/K_F$ (5), which contains no legal primitive; the illegal price is $c_F + \mu$ (Lemma 1(i)). (ii) Each legal share solves $g(s_\ell) = (m_\ell - \mu)/\mu - \ln \bar{A}$ (Lemma 1(ii)), and $m_\ell = v_\ell - c_\ell - \tau$ falls one for one with τ, so every share falls (Lemma 1(iii)) and S_L falls. Since $S_F = 1 - 1/\bar{A} - S_L$, $dS_F = -dS_L$. Each illegal seller's share is $\widehat{r}_f/\bar{A} = K_F/\mu$, so $N^* = (\mu/K_F)S_F$. Since $1 + A_L + A_F = \bar{A}$, $dA_F = -dA_L$. (iii) Differentiating the share equation, with $g'(s) = 1/s + 1/(1-s)^2$, gives $d\ln s_\ell/d\tau = -\kappa_\ell/\mu$; with $p_\ell = c_\ell + \tau + \mu/(1-s_\ell)$, $dp_\ell/d\tau = 1 - \kappa_\ell s_\ell/(1-s_\ell)^2 = \kappa_\ell$. □

Because the break-even aggregate $\bar{A}$ in part (i) contains no legal primitive, a tax cannot change the size of the market while illegal sellers are active; it changes only the composition of sales. Legal entry has the same effect.

Corollary 1 (Legal entry and license caps)**.** *Hold τ fixed and suppose illegal sellers are active before and after a legal firm k is added. Then $\bar{A}$, consumer surplus, the illegal price, and every*

incumbent legal firm's price, share and profit are unchanged; illegal entry falls by $\widetilde{r}_k(\bar{A})/\widehat{r}_f$*, and illegal sales fall and legal sales rise by k's share* $s_k = \widetilde{r}_k(\bar{A})/\bar{A}$.

Proof. Both legal sets leave illegal sellers active, so both continuations have $A^* = \bar{A}$ (9), and each incumbent plays $\widetilde{r}_\ell(\bar{A})$ in both, with the same share, price and profit (Lemma 1). By (8), N^* falls by $\widetilde{r}_k(\bar{A})/\widehat{r}_f$. S_L rises by s_k, so $S_F = 1 - 1/\bar{A} - S_L$ falls by s_k. □

Under logit, a license cap does not raise incumbents' markups; it raises illegal sales, and with them illegal entry, one for one with the legal sales it prevents. This neutrality toward incumbents is specific to logit (Remark 3). The licensed set is held fixed in the estimation and the counterfactuals.[15]

A.1.2 Legal Leadership

In the leader game, legal firms commit to prices before illegal sellers enter. The leader game is not the estimated model, whose legal markup is the simultaneous one (1); it is the benchmark from the leadership literature and shows which results depend on timing.

Etro (2006, 2008) study leadership with free entry of followers: leaders are more aggressive and, under price competition, price below the followers, the reverse of the fixed-entry result. With several leaders and free follower entry, Alfaro and Lander (2021) find that all leaders restrict entry without harming one another, and Ino and Matsumura (2012) find that quantity leadership raises welfare over Cournot. Multi-leader games more generally are surveyed by Hu and Fukushima (2015), with existence results in Mallozzi and Messalli (2017) and Aussel and Svensson (2018), and a separate literature studies a single dominant firm facing a competitive fringe (White, 1981; Kahai et al., 1996; Golombek et al., 2018). Here several heterogeneous legal firms, whose entry is endogenous, choose prices before a free-entry fringe of illegal sellers enters, and the equilibrium is unique.

Timing and the leader's reply. The timing is: (1) legal candidates enter in rank order and then set prices simultaneously; (2) illegal sellers observe legal prices, enter, and set prices; (3) consumers choose a legal product, an illegal product, or the outside good. Legal firms are dominant firms that set prices before a competitive fringe enters, as in the Stackelberg and dominant-firm models of Etro (2008); Anderson et al. (2020); Forchheimer (1908).

After any legal actions **a**, illegal entry and the aggregate are given by Lemma 5. While illegal sellers are active the aggregate stays at $\bar{A}$, so a legal firm perceives $\partial A/\partial a_\ell = 0$ and plays the

[15] Washington's excise tax is ad valorem. Let τ_{st} be the tax wedge of Section 4, the share of the customer's payment that goes to taxes. Legal operating profit is $(1-\tau_{st})\big(p_\ell - c_\ell/(1-\tau_{st})\big)s_\ell$, so the game is the per-unit game with tax-inclusive cost $c_\ell/(1-\tau_{st})$ and fixed cost $K_\ell/(1-\tau_{st})$. Lemma 4(i)–(ii) and Corollary 1 carry over unchanged, and pass-through becomes $dp_\ell/d\tau_{st} = \kappa_\ell c_\ell/(1-\tau_{st})^2$.

fixed-aggregate reply

$$\widehat{r}_\ell : \quad \frac{\partial \pi_\ell(a_\ell, A)}{\partial a_\ell} + \frac{\partial \pi_\ell(a_\ell, A)}{\partial A}\frac{\partial A}{\partial a_\ell} = \frac{\partial \pi_\ell(a_\ell, A)}{\partial a_\ell} = 0, \tag{44}$$

which under logit is the constant $\widehat{r}_\ell = e^{(m_\ell - \mu)/\mu}$, with markup μ. Because free entry holds the aggregate fixed, legal leaders deter part of illegal entry by taking a larger action than in the simultaneous game, $\widehat{r}_\ell > \widetilde{r}_\ell(\bar{A})$, as leaders facing free entry do in Etro (2006, 2008), and a larger action than leaders facing a fixed number of followers, for whom $dA/da_\ell > 1$. Only the leaders' own first-order condition changes: the aggregate and each illegal seller's action are the same as in the simultaneous game, and illegal entry is lower. Enough legal leaders can exclude illegal sellers altogether by limit pricing.

Equilibrium.

Lemma 5 (Follower response and the leader's payoff)**.** *Fix a nonempty prefix P_L. Followers enter after observing legal actions* **a***, so*

$$N(\mathbf{a}) = \max\left\{0,\ \frac{\bar{A} - 1 - \sum_\ell a_\ell}{\widehat{r}_f}\right\}, \qquad A(\mathbf{a}) = \max\left\{\bar{A},\ 1 + \sum_\ell a_\ell\right\}.$$

With rivals fixed and $\bar{b}_\ell = \bar{A} - 1 - \sum_{k\neq\ell} a_k$, leader ℓ's operating payoff from action b is $\pi_\ell(b) = (m_\ell - \mu \ln b)\, b/\bar{A}$ for $b \le \bar{b}_\ell$ (fringe active) and $(m_\ell - \mu \ln b)\, b/(1 + \sum_{k\neq\ell} a_k + b)$ for $b \ge \bar{b}_\ell$ (no fringe).

(i) $\pi_\ell(b) \le \mu\widehat{r}_\ell/\bar{A}$, the payoff from $\widehat{r}_\ell$ at $\bar{A}$.

(ii) The left piece is strictly concave with peak $\widehat{r}_\ell$; the right piece is single-peaked (A2).

(iii) At a profile with $1 + \sum_\ell a_\ell = \bar{A}$, $\pi'_\ell(a_\ell^-) \ge 0$ if and only if $a_\ell \le \widehat{r}_\ell$, and $\pi'_\ell(a_\ell^+) \le 0$ if and only if $a_\ell \ge \widetilde{r}_\ell(\bar{A})$.

Proof. (i) $A(\mathbf{a}) \ge \bar{A}$, so a positive payoff is at most $(m_\ell - \mu \ln b)b/\bar{A} \le \mu\widehat{r}_\ell/\bar{A}$ by (ii). (ii) $(m_\ell - \mu \ln b)b$ has derivative $m_\ell - \mu - \mu \ln b$, zero at $\widehat{r}_\ell$, and second derivative $-\mu/b < 0$. (iii) Here $\bar{A}\pi'_\ell(a_\ell^-) = m_\ell - \mu \ln a_\ell - \mu$ and $\bar{A}\pi'_\ell(a_\ell^+) = (m_\ell - \mu \ln a_\ell)(1 - a_\ell/\bar{A}) - \mu$; the second is nonpositive if and only if $\ln(\widehat{r}_\ell/a_\ell) \le a_\ell/(\bar{A} - a_\ell)$, whose left side falls and right side rises in a_ℓ, with equality at $\widetilde{r}_\ell(\bar{A})$ (Lemma 1(ii)). □

By A4, $(1 + \sum_\ell \widetilde{r}_\ell(A))/A$ is strictly decreasing and equals one at $\widetilde{A}(P_L)$, so $1 + \sum_\ell \widetilde{r}_\ell(\bar{A}) \le \bar{A}$ if and only if $\widetilde{A}(P_L) \le \bar{A}$; and $\widetilde{A}(P_L) < 1 + \sum_\ell \widehat{r}_\ell$ because $\widetilde{r}_\ell < \widehat{r}_\ell$. Hence every nonempty prefix

lies in exactly one of three regions:

$$\text{(i)}\ \ 1+\sum_{\ell\in P_L}\widehat{r}_\ell<\bar{A};\qquad \text{(ii)}\ \ \widetilde{A}(P_L)\le\bar{A}\le 1+\sum_{\ell\in P_L}\widehat{r}_\ell;\qquad \text{(iii)}\ \ \widetilde{A}(P_L)>\bar{A}.$$

Proposition 1L. ***Multiple-Leader Equilibrium and Global Uniqueness.*** *Under logit demand and A5, legal entry, legal price commitment, and subsequent illegal entry have a unique selected subgame-perfect equilibrium. For every nonempty prefix the continuation lies in one of the following regions, where $\widetilde{A}=\widetilde{A}(P_L)$:*

(i) Accommodation. $1+\sum_{\ell\in P_L}\widehat{r}_\ell<\bar{A}$. *Legal actions are $\widehat{r}_\ell$, $A^*=\bar{A}$, and $N^*=(\bar{A}-1-\sum\widehat{r}_\ell)/\widehat{r}_f>0$.*

(ii) Limit pricing. $\widetilde{A}\le\bar{A}\le 1+\sum\widehat{r}_\ell$. *The pricing equilibrium set is*

$$\mathcal{E}(P_L)=\left\{\mathbf{a}:\ \sum a_\ell=\bar{A}-1,\quad \widetilde{r}_\ell(\bar{A})\le a_\ell\le\widehat{r}_\ell\right\}.$$

Here $N^*=0$ and $A^*=\bar{A}$. Deterrence priority A5(b) gives

$$a_\ell^*=\min\{\widehat{r}_\ell,R_\ell\},\qquad R_\ell=\bar{A}-1-\sum_{k<\ell}a_k^*-\sum_{k>\ell}\widetilde{r}_k(\bar{A}).$$

Without A5(b) the set is a continuum if $L\ge 2$ and both region inequalities are strict; every profile in it has the same aggregate, mass and consumer surplus.

(iii) Ordinary exclusion. $\widetilde{A}>\bar{A}$. *Legal actions are $\widetilde{r}_\ell(\widetilde{A})$, $A^*=\widetilde{A}$, and $N^*=0$.*

For every prefix $A^*(P_L)=\max\{\bar{A},\widetilde{A}(P_L)\}$; (D1)–(D2) hold at the selected continuations. The legal set is given by the same maximum-profitable-prefix formula and first-failure characterization as in Proposition 1S, with leadership profits in Γ_L.

Proof. Write $\widetilde{A}=\widetilde{A}(P_L)$. By the paragraph before the proposition, regions (i)–(iii) are exhaustive and disjoint, and region (ii) is exactly $\sum_\ell\widetilde{r}_\ell(\bar{A})\le\bar{A}-1\le\sum_\ell\widehat{r}_\ell$. By Lemma 5, a leader earns $h_\ell(a_\ell)/\bar{A}$ below the kink, where $h_\ell(a)=(m_\ell-\mu\ln a)a$ peaks at $\widehat{r}_\ell$, and its ordinary payoff above it. An equilibrium below the kink is therefore $\widehat{\mathbf{r}}=(\widehat{r}_\ell)$, one at the kink has $\widetilde{r}_\ell(\bar{A})\le a_\ell\le\widehat{r}_\ell$, and one above it is the ordinary profile, with total $\widetilde{A}-1$.

Case 1: Legal firms and illegal sellers enter (accommodation). In (i), $1+\sum\widehat{r}_\ell<\bar{A}$, so $\widehat{\mathbf{r}}$ lies below the kink and each leader attains its bound $\mu\widehat{r}_\ell/\bar{A}$. A kink profile needs $\bar{A}-1\le\sum\widehat{r}_\ell$, which fails, and the ordinary profile lies below the kink but differs from $\widehat{\mathbf{r}}$. So $\widehat{\mathbf{r}}$ is the only equilibrium. For a given legal set, the aggregate, consumer surplus, and illegal price are those of Case 1 of

Proposition 1S; since $\widehat{r}_\ell > \widetilde{r}_\ell(\bar{A})$, legal shares are larger and illegal entry, $N^* = (\bar{A} - 1 - \sum \widehat{r}_\ell)/\widehat{r}_f$, is lower.

Case 2a: Legal firms enter and illegal sellers do not (ordinary exclusion). In (iii), $\widetilde{A} > \bar{A}$, so only the ordinary profile is a candidate. Write $D_\ell = 1 + \sum_{k \neq \ell} a_k$, so that $\bar{b}_\ell = \bar{A} - D_\ell$ in Lemma 5. A deviation $b < a_\ell$ that invites followers keeps a positive margin and earns

$$\frac{h_\ell(b)}{\bar{A}} < \frac{h_\ell(b)}{D_\ell + b} \leq \pi_\ell(a_\ell),$$

because $D_\ell + b < \bar{A}$ and the no-follower payoff $h_\ell(b)/(D_\ell + b)$ is single-peaked on all $b > 0$ with peak a_ℓ (Lemma 1(ii)). The ordinary profile is therefore the equilibrium. Ordinary prices already leave no room for followers, and small deviations keep the total above the kink, so commitment changes nothing: the outcome is that of Case 2 of Proposition 1S, with $N^* = 0$ and $A^* = \widetilde{A} > \bar{A}$ set by legal primitives.

Case 2b: Legal firms limit price and illegal sellers stay out. Region (ii) is exactly the condition for the kink set $\mathcal{E}(P_L)$ to be nonempty, and single-peakedness makes its tests global. No equilibrium lies off the kink: $\sum \widehat{r}_\ell \geq \bar{A} - 1$ keeps $\widehat{\mathbf{r}}$ from lying below it, and $\widetilde{A} \leq \bar{A}$ keeps the ordinary profile from lying above it. On $\mathcal{E}(P_L)$ each profit $h_\ell(a_\ell)/\bar{A}$ rises in a_ℓ up to $\widehat{r}_\ell$, so A5(b) gives each firm in turn the largest action that later firms' lower bounds allow. This recursion defines a^*_ℓ. It is unique, so action ties never arise, and it lies in $\mathcal{E}(P_L)$: $R_1 \geq \widetilde{r}_1(\bar{A})$ because $\sum_\ell \widetilde{r}_\ell(\bar{A}) \leq \bar{A} - 1$, and $R_{\ell+1} = R_\ell - a^*_\ell + \widetilde{r}_{\ell+1}(\bar{A}) \geq \widetilde{r}_{\ell+1}(\bar{A})$, so every action meets its lower bound; once a firm receives R_ℓ, every later firm receives its lower bound and the total is $\bar{A} - 1$; and some firm receives its residual, since otherwise $\sum_\ell \widehat{r}_\ell < \bar{A} - 1$. Leaders hold $1 + \sum a_\ell$ exactly at $\bar{A}$, where a first follower would just break even, so $N^* = 0$ while A^* and consumer surplus equal their active-fringe values: the threat of entry alone disciplines prices. Commitment creates the kink in each payoff (Lemma 5(iii)), which is why every split of $\bar{A} - 1$ between the bounds is an equilibrium; A5(b) only divides profit among leaders and leaves A^*, N^*, and consumer surplus unchanged.

Legal ranking. Selected actions fall with the index: in (i) and (iii) by the margin order, and in (ii) because an earlier firm gets its larger upper bound or a residual after which later firms get their lower bounds. Every action lies at or below its peak, where h_ℓ rises in both a and m_ℓ, so profits fall with the index and $K_1 \leq \cdots \leq K_L$ gives (D1). Adding firm $L + 1$ weakly raises A^* and moves the prefix to the same or a later region, and

$$\pi^*_{L+1} < \begin{cases} \bar{\pi}_L(A^*(P_{L+1})) \leq \bar{\pi}_L(A^*(P_L)), & P_{L+1} \text{ in (iii)}, \\ \mu \widehat{r}_L/\bar{A}, & P_L \text{ in (i)}, \\ \bar{\pi}_L(\bar{A}), & P_L, P_{L+1} \text{ in (ii)}. \end{cases}$$

The first row is Lemma 1(iv) and the second Lemma 5(i) with $\widehat{r}_{L+1} < \widehat{r}_L$. In the last case firm $L+1$ receives its lower bound $\widetilde{r}_{L+1}(\bar{A})$: because $\sum_{k\le L}\widehat{r}_k \ge \bar{A}-1$, one of firms $1,\dots,L$ receives its residual. Each bound is at most $\pi^*_L(P_L)$: in (iii) $\pi^*_L(P_L)=\bar{\pi}_L(A^*(P_L))$, and in (i)–(ii) $\widetilde{r}_L(\bar{A}) \le a^*_L \le \widehat{r}_L$ and h_L rises up to $\widehat{r}_L$. Selected profits are positive, so dividing by $K_{L+1} \ge K_L$ gives (D2).

Case 3: Illegal sellers enter and legal firms do not ($L^*=0$, $\bar{A}>1$): the empty prefix gives $A^*=\bar{A}$ and $N^*=(\bar{A}-1)/\widehat{r}_f$, as in Proposition 1S. With no legal firm there is nothing to commit to, so the timing does not matter. No candidate enters because $\Gamma_L \le 1$ for every L, now with leadership profits, which under (D2) reduces to $\Gamma_1 \le 1$.

Case 4: Neither enters ($L^*=0$, $\bar{A}\le 1$): $A^*=1$ and $N^*=0$. As in Proposition 1S, a seller would earn $\mu\widehat{r}_f \le K_F$ even against the outside good alone, and no candidate covers its fixed cost, so only the outside good is bought.

Every prefix therefore has one selected continuation, and Lemma 2 gives L^*. □

Comparative statics and welfare.

Remark 2 (Neutrality with illegal sellers active). In the leader game, suppose illegal sellers are active before and after the change. A change in one legal firm's cost or tax, the entry of a legal firm, or a merger of legal firms that keeps all products leaves A, consumer surplus and every other firm's action unchanged; N falls by the rise in $\sum_\ell a_\ell$ divided by $\widehat{r}_f$; and the change in welfare equals the change in the affected firms' profit plus tax revenue, minus the change in harms. Zero profit holds the aggregate at $\bar{A}$ (Lemma 5), each leader plays $\widehat{r}_\ell$, which depends only on its own margin (a merger leaves each product's fixed-aggregate condition unchanged, since $dA/da_j=0$), market clearing gives N, and consumer surplus is $\mu\ln\bar{A}$ throughout. In Case 1 of the simultaneous game, by contrast, a merger of two legal firms that keeps both products raises the merged products' prices at $\bar{A}$ and raises N, leaving consumer surplus unchanged: the merged firm equates the markups on its two products and so acts as one legal firm with $\widehat{r}_{jk}=\widehat{r}_j+\widehat{r}_k$, hence a larger share and markup (Lemma 1(ii)); other legal firms keep $\widetilde{r}_\ell(\bar{A})$, so $\sum_\ell a_\ell$ falls and N rises (Lemma 3(i)).

A.2 Sub-Aggregative Games: Nested Logit

This section extends the analysis to nested logit. The results are similar, but the proofs are longer, and the assumptions B1–B6 are stated below. Demand is nested logit with a legal nest of L firms, an illegal nest with a mass N of sellers, and the outside good. Nested demand with free entry is studied by Anderson and de Palma (1992, 2006), where each nest is the product range of one of many symmetric firms, and by Anderson et al. (2015), who extend free entry to nested aggregative games with insiders and entrants in each nest. Here the nests are the legal and illegal sectors, and

only the illegal nest has free entry.

Legal products form nest L and illegal sellers nest F; the outside good is in neither nest. Nest scales are $\mu_J = \rho_J \mu$ with $\rho_L \in (0, 1]$ and $\rho_F \in (0, 1)$. Actions are $a_j = e^{(v_j - p_j)/\mu_J}$ for j in nest J, nest sub-aggregates are $A_L = \sum_{\ell \in P} a_\ell$ and $A_F = N a_f$, and the aggregate is $A = 1 + A_L^{\rho_L} + A_F^{\rho_F}$. Shares are $s_0 = 1/A$, $S_J = A_J^{\rho_J}/A$, and $s_j = s_{j|J} S_J$ with $s_{j|J} = a_j / A_J$. Operating profits are

$$\pi_\ell = (m_\ell - \mu_L \ln a_\ell) \frac{a_\ell}{A_L} \frac{A_L^{\rho_L}}{A}, \qquad \pi_f = (v_F - c_F - \mu_F \ln a_f) \frac{a_f}{A_F} \frac{A_F^{\rho_F}}{A}.$$

Write $X \equiv A_L^{\rho_L}$ for the legal inclusive value, $\sigma \equiv 1 - \rho_F$ and $\gamma \equiv \rho_F / (1 - \rho_F)$.

Each illegal seller plays $a_f^0 = \exp((v_F - c_F - \mu_F)/\mu_F)$, at price $c_F + \mu_F$, so $A_F = N a_f^0$ (Lemma 6(i)); its operating profit per unit of mass is $\widehat{\pi}_F(A_F, A) = \mu_F a_f^0 A_F^{\rho_F - 1}/A$, and $H \equiv \mu_F a_f^0 / K_F$. A legal firm that sets its price with the mass of illegal sellers fixed has markup $\mu_L/(1 - \chi s_{\ell|L})$, $\chi \equiv 1 - \rho_L + \rho_L S_L$ (Lemma 6(iii)). With no illegal sellers, $\chi_0 = 1 - \rho_L + \rho_L A_L^{\rho_L}/(1 + A_L^{\rho_L})$, and $\widetilde{r}_{\ell,L}^0(A_L)$ is the ordinary nested reply at legal sub-aggregate A_L.

Assumptions.

- **B1: Competitiveness.** For every firm in nest J,

  $$\frac{\partial \pi_{j,J}}{\partial A} < 0, \qquad \frac{\partial \pi_{j,J}}{\partial A_J} \leq 0$$

  at a positive margin.

- **B2: Unique Own-Action Optimum.** Profits $\pi_{j,J}$ are strictly quasi-concave in $a_{j,J}$, so the relevant first-order condition delivers a unique maximum.

- **B3: Nested Strategic Complementarity.** On each smooth strict branch, the relevant legal inclusive best replies are continuous and positive. On the no-follower branch, ordinary legal inclusive best replies are increasing in the legal sub-aggregate and the aggregate.

- **B4: Nested Fixed-Set Slope Condition.** For each candidate legal set P_L and each strict branch $s \in \{0, F\}$, define the normalized legal reply sum

  $$G_L^s(A_L) = \frac{\sum_{\ell \in P_L} \widetilde{r}_{\ell,L}^s(A_L)}{A_L}.$$

  Each G_L^s is continuous and strictly decreasing, so it crosses one at most once.

  Under nested logit most of B1–B4 follow from the primitives. At a positive margin $\partial \pi_{j,J}/\partial A < 0$ and $\partial \pi_{j,J}/\partial A_J = (\rho_J - 1)\pi_{j,J}/A_J \leq 0$, with equality in the legal nest at

$\rho_L = 1$; no proof below uses B1 or B3. Lemma 6(iii) proves B2 at every fixed follower mass and B4 on the no-follower branch, so Proposition 3S uses none of B1–B4 as an assumption. The leader game uses only B2, for the payoff on the active branch, where followers respond along the zero-profit locus (Lemma 7(i)).

- **B5: Rank-Ordered Legal Entry and Priority.** Legal candidates are indexed in decreasing efficiency order and only sets $P_L = \{1, \dots, L\}$ are feasible. For each set, let $\pi_L^*(P_L)$ be the marginal legal firm's equilibrium operating profit in its unique (selected) nested continuation, and $\Gamma_L^N = \pi_L^*(P_L)/K_L$.

  (a) **Entry priority.** As A5(a), with Γ^N in place of Γ.

  (b) **Deterrence priority.** As A5(b), with Γ^N in place of Γ. Under B6, B5(b) applies to the whole equilibrium set of the leader pricing stage: the selected deterrence profile is compared with accommodation by the same rank rule, and exact ties go to deterrence. Without B6 it is unnecessary, because no continuation deters the fringe.

  (c) **Ordering.** As A5(c): $v_1 - c_1 > v_2 - c_2 > \cdots > v_{\bar{L}} - c_{\bar{L}}$ and $K_1 \le K_2 \le \cdots \le K_{\bar{L}}$. Unlike under logit, Γ^N need not fall with L; the legal-set formula uses only (D1) (Remark 1).

- **B6: Minimum Illegal Mass.** The illegal sector is either inactive or has at least a minimum viable mass:

$$N \in \{0\} \cup [\underline{N}, \infty), \qquad \underline{N} > 0.$$

  Conditional on activation, every individual illegal seller remains atomistic. A minimum mass that would exactly break even stays out.

Lemma 6 (Nested logit continuation)**.** *Under the nested primitives above:*

(i) Free-entry locus. *Each active illegal seller plays a_f^0 and charges $c_F + \mu_F$. For every $A_L \ge 0$ there is exactly one positive A_F satisfying the aggregate identity $A = 1 + X + A_F^{\rho_F}$ and*

$$A\, A_F^{1-\rho_F} = H.$$

Along that identity, illegal-seller profit falls strictly with A_F and with A_L. On the locus,

$$A_F^{\rho_F} = H^{\gamma} A^{-\gamma}, \qquad S_F = A_F/H, \qquad \frac{dA}{dX} = \frac{1}{1 + \gamma S_F} \in (0, 1).$$

As A_L rises, S_L rises and A_F, S_F fall.

(ii) Minimum mass and the activation cutoff. *Under B6, write $\underline{A}_F = \underline{N} a_f^0$ and $(A_L^{\dagger})^{\rho_L} = \max\{0,\ H \underline{A}_F^{\rho_F - 1} - 1 - \underline{A}_F^{\rho_F}\}$. At fixed legal actions a minimum mass profits exactly when*

$A_L < A_L^\dagger$. *In that case the unique continuation lies on the locus in (i), with* $N > \underline{N}$*; otherwise* $N = 0$. *If* $A_L^\dagger > 0$*, the active-side aggregate limit at* $A_L = A_L^\dagger$ *is* $1 + (A_L^\dagger)^{\rho_L} + \underline{A}_F^{\rho_F}$*, whereas its actual cutoff value is* $1 + (A_L^\dagger)^{\rho_L}$.

(iii) Fixed-mass prices and post-entry activation. *For every prefix* P_L *and fixed* $N \geq 0$*, the legal pricing equilibrium is unique, with sub-aggregate* $A_L(N)$*, where* $A_L(N) = 0$ *for* P_0. *For a nonempty prefix it is continuous and strictly increasing in* N *and each legal margin, as is* $A(N) = 1 + A_L(N)^{\rho_L} + (Na_f^0)^{\rho_F}$. *Legal markups are*

$$p_\ell - c_\ell - \tau = \frac{\mu_L}{1 - \chi s_{\ell|L}}, \qquad \chi = 1 - \rho_L + \rho_L S_L.$$

Post-entry illegal-seller profit $\Pi_F(N) = \widehat{\pi}_F(Na_f^0, A(N))$ *falls continuously and strictly from* $+\infty$ *to zero on* $N > 0$. *Under B6, a minimum mass enters exactly when* $\Pi_F(\underline{N}) > K_F$*, equivalently* $A_L(\underline{N}) < A_L^\dagger$. *The no-follower profile has total* $A_L(0)$*, the legal-only fixed point in B4.*

Proof. *(i)* Holding A_F, A fixed, an illegal seller maximizes a positive multiple of $(m_F - \mu_F \ln a_f)a_f$. Strict concavity gives a_f^0 and its price. The zero-profit locus $A = HA_F^{\rho_F - 1}$ falls from infinity to zero, while the identity $A = 1 + X + A_F^{\rho_F}$ rises from $1 + X$ to infinity; they cross once. Along the identity,

$$\frac{\partial \ln \widehat{\pi}_F}{\partial A_F} = \frac{\rho_F - 1}{A_F} - \frac{\rho_F A_F^{\rho_F - 1}}{A} < 0.$$

Increasing A_L lowers profit through A. On the locus, $A = 1 + X + H^\gamma A^{-\gamma}$, so $(1 + \gamma S_F)dA = dX$. The displayed identities follow; since A rises, s_0, S_F fall and $S_L = 1 - s_0 - S_F$ rises.

(ii) At the minimum mass, profit exceeds K_F exactly when $A_L^{\rho_L} < H\underline{A}_F^{\rho_F - 1} - 1 - \underline{A}_F^{\rho_F}$. If it does, the decreasing profit curve crosses K_F at a larger mass; if not, every larger network also fails the activation test. Equality is assigned to no entry. At a positive cutoff the locus approaches $A_F = \underline{A}_F$, giving the stated one-sided aggregate limit.

(iii) With rivals and N fixed, write $a = a_\ell$ and $M = m_\ell - \mu_L \ln a$. On $M > 0$,

$$\frac{\partial \ln \pi_\ell}{\partial \ln a} = 1 - \chi s_{\ell|L} - \frac{\mu_L}{M}. \tag{45}$$

Both χ and $s_{\ell|L}$ rise with a, so the derivative decreases strictly from a positive limit to $-\infty$. This includes a singleton: its lower-end limit is $\rho_L > 0$. At nonpositive margins the payoff is nonpositive and decreasing, since its share factor rises. Thus the unique smooth peak is global. At a candidate

total $T = A_L > 0$, the formal share solves

$$\ln(Ts) + \frac{1}{1 - \chi(T;N)s} = \frac{m_\ell}{\mu_L}, \qquad 0 < s < 1/\chi(T;N). \tag{46}$$

The left side rises from $-\infty$ to $+\infty$ in s. Each formal share falls with T, rises with N (which lowers χ), and rises with its margin. As $T \to 0$, each formal share exceeds one; as $T \to \infty$, it tends to zero. Their sum therefore crosses one exactly once, giving the unique pricing profile and the continuity and monotonicity of its total. The actual equilibrium shares are in $(0, 1]$; formal shares above one are used only for the endpoint existence argument.

In $\Pi_F(N)$ the numerator $(Na_f^0)^{\rho_F - 1}$ falls from infinity to zero and the denominator rises from the finite positive $1 + A_L(0)^{\rho_L}$. This gives the profit limits and strict monotonicity. At $N = \underline{N}$, part (ii) evaluated at post-entry prices gives the activation test. □

Proposition 3S. ***Nested Simultaneous Equilibrium and Global Uniqueness.*** *Under the nested primitives, B5(a), B5(c), and B6, legal entry, illegal entry, and simultaneous pricing have a unique subgame-perfect equilibrium. Accommodation and exclusion cannot coexist for a fixed prefix, and no deterrence priority is needed. Within-prefix ranking (D1) holds, and*

$$L^* = \max\big(\{0\} \cup \{L \in \{1, \dots, \bar{L}\} : \Gamma_L^N > 1\}\big).$$

Every entrant's final ratio exceeds one, and $\Gamma_L^N \le 1$ for $L > L^*$. Without B6 the result continues to hold and the fringe is always active, with the unique positive zero-profit mass.

Proof. Fix a prefix. By Lemma 6(iii), each mass N has one pricing equilibrium, with legal shares solving (46), and post-entry illegal-seller profit

$$\Pi_F(N) = \frac{\mu_F a_f^0 \, (Na_f^0)^{\rho_F - 1}}{A(N)}$$

falls strictly from $+\infty$ to 0. Hence $N^* > \underline{N}$ is the unique root of $\Pi_F(N^*) = K_F$ when $\Pi_F(\underline{N}) > K_F$, that is, when $A_L(\underline{N}) < A_L^\dagger$, and $N^* = 0$ otherwise. Each prefix has one continuation, so accommodation and exclusion cannot coexist and no selection is needed. Firms in a prefix share S_L and χ, so (46) orders $s_{j|L}$ by margin, and operating profit $\mu_L S_L s_{j|L}/(1 - \chi s_{j|L})$ rises with $s_{j|L}$; with B5(c) this is (D1), and Lemma 2 selects P_{L^*}. In every case $A^* = 1 + (A_L^*)^{\rho_L} + (A_F^*)^{\rho_F}$, and legal markups are $\mu_L/(1 - \chi s_{\ell|L})$.

Case 1: Legal firms and illegal sellers enter ($L^* > 0$, $A_L(\underline{N}) < A_L^\dagger$):

$$\Pi_F(N^*) = K_F, \qquad A_L^* = A_L(N^*), \qquad A_F^* = N^* a_f^0. \tag{47}$$

Unlike logit, zero profit no longer pins A; it pins the locus $A\,A_F^{1-\rho_F} = H$ of Lemma 6(i), so legal prices change both the size and the composition of the aggregate. Legal firms price with N fixed and internalize their full effect on A, through $\chi = 1 - \rho_L + \rho_L S_L$. Locally, of each sale the legal sector loses, a fraction $S_F/(S_F + \sigma s_0)$ goes to illegal sellers and the rest leaves the market, as in Corollary 2(iii).

Case 2: Legal firms enter and illegal sellers do not ($L^* > 0$, $A_L(\underline{N}) \geq A_L^\dagger$): $N^* = 0$ and $A_L^* = A_L(0)$. Either $A_L(0) \geq A_L^\dagger$, or $A_L(0) < A_L^\dagger \leq A_L(\underline{N})$. In the second case a minimum mass would profit at the no-follower prices but not at the lower legal prices that follow its entry, since $A_L(\underline{N}) > A_L(0)$, so legal firms exclude without committing to limit prices. In both parts legal firms play the ordinary nested profile, with χ_0, and $A^* = 1 + A_L(0)^{\rho_L}$ is set by legal primitives alone. This case exists only because of B6: without a minimum mass the first entrants would earn unbounded profit.

Case 3: Illegal sellers enter and legal firms do not ($L^* = 0$, $A_L^\dagger > 0$): at $A_L = 0$ a minimum mass profits, and $A_F^* > \underline{A}_F$ solves $(1 + A_F^{\rho_F})A_F^{1-\rho_F} = H$. This is the locus of Lemma 6(i) with $A = 1 + A_F^{\rho_F}$, and $A_L^\dagger > 0$ is the activation test of Lemma 6(ii) at $A_L = 0$; illegal primitives alone set the aggregate and $N^* = A_F^*/a_f^0$. No candidate enters because $\Gamma_L^N \leq 1$ for every L. Unlike logit, this cannot be read off Γ_1^N, because (D2) is not available (Remark 3).

Case 4: Neither enters ($L^* = 0$, $A_L^\dagger = 0$): $A_L^* = A_F^* = N^* = 0$ and $A^* = 1$. Here $A_L^\dagger = 0$ means $H\underline{A}_F^{\rho_F - 1} \leq 1 + \underline{A}_F^{\rho_F}$: a minimum mass loses even against the outside good alone, and by Lemma 6(i) every larger one loses more. With $\Gamma_L^N \leq 1$ for every L, only the outside good is bought.

Without B6. Π_F still falls from $+\infty$ to 0, so every prefix, including the empty one, has a unique $N^* > 0$; only Cases 1 and 3 remain. □

Corollary 2 (Nested logit without minimum entry)**.** *Without B6, as in the estimated model, and holding illegal primitives fixed:*

(i) Prices satisfy $p_f = c_F + \mu_F$ and $p_\ell = c_\ell + \tau + \mu_L/(1 - \chi s_{\ell|L})$, $\chi = 1 - \rho_L + \rho_L S_L$.

(ii) The fringe is active in every continuation; its zero-profit locus is equivalently $S_F = H^\gamma s_0^{1/\sigma}$.

(iii) For a local change in the legal sector along that locus with $dS_L \neq 0$,

$$-\frac{dS_F}{dS_L} = \frac{S_F}{S_F + \sigma s_0}, \qquad \frac{ds_0}{-dS_L} = \frac{\sigma s_0}{S_F + \sigma s_0}.$$

(iv) For a fixed nonempty legal set, a higher tax lowers S_L and raises s_0, S_F, N, with $d\ln N = d\ln S_F$.

Proof. Parts (i)–(ii) follow from Lemma 6 and Proposition 3S. With $\rho_L = \rho_F = 1 - \sigma$ and $\mu = 1/\alpha$, $\chi s_{\ell|L} = \sigma s_{\ell|L} + (1 - \sigma)s_\ell$, so the markup in (i) is the nested-logit markup of Berry (1994). Since

$1+\gamma = 1/\sigma$, the locus gives $S_F = H^\gamma A^{-(1+\gamma)} = H^\gamma s_0^{1/\sigma}$, with the converse obtained by reversing this algebra. Thus $d\ln S_F = \sigma^{-1} d\ln s_0$; combine it with $-dS_L = dS_F + ds_0$ to obtain (iii). At fixed N, a tax lowers legal margins, hence $A_L(N)$ and $A(N)$, raising $\Pi_F(N)$. Its decreasing curve therefore reaches K_F at a larger mass. On the locus $A = HA_F^{\rho_F - 1}$ falls as $A_F = Na_f^0$ rises, so s_0, S_F rise and S_L falls. Finally $S_F = Na_f^0/H$ gives the log-change identity. □

Proposition 3S and Corollary 2(i)–(ii) show that the structure of Section 3.4 survives, with two changes. First, without minimum illegal entry, illegal sellers are active in every equilibrium, because as the illegal nest shrinks each remaining seller's profit grows without bound. Second, the zero-profit condition no longer pins the aggregate; it pins a locus on which the illegal share moves with the outside share, so Corollary 2(iii) replaces Lemma 4(ii): of each sale the legal sector loses, a fraction $S_F/(S_F + \sigma s_0)$ goes to illegal sellers and the rest leaves the market. At given shares, free entry raises this fraction above $S_F/(S_F + s_0)$, its value when the number of illegal sellers is held fixed, and the increase is smaller the closer σ is to one.

For the estimated model the theory implies uniqueness for the observed legal set, an always-active fringe, a constant illegal price, the markups in Corollary 2(i) and the signs in Corollary 2(iv). It does not imply a fixed market size or constant consumer surplus, the pass-through formula of Lemma 4(iii), or incumbent-neutral legal entry (Corollary 1). Whether Γ_L^N falls with L is not proved (Remark 3); the counterfactuals hold the legal set fixed, so this does not affect them.

Remark 3 (Legal entry under nested logit). Within-prefix ranking is sufficient for the maximum-profitable-prefix formula. A legal firm's profit depends on the legal set even with an active fringe, and the logit proof of (D2) does not carry over: changes in nest shares and illegal entry can offset the direct competitive effect of an added firm. Neither an incumbent-profit comparison nor the primitives above establishes a decreasing marginal-entry ratio. No (D2) assumption is used in Proposition 3S or 3L.

A.2.1 Nested Leadership

Lemma 7 (Leader pricing stage under B6)**.** *Fix a nonempty prefix P_L under B6. Followers observe legal prices: by Lemma 6(ii) they enter, with aggregate $A^F(A_L)$ on the zero-profit locus, if $A_L < A_L^\dagger$, and stay out otherwise. Leader ℓ's operating profit is therefore*

$$\pi_\ell^F(\mathbf{a}) = \frac{(m_\ell - \mu_L \ln a_\ell)\, a_\ell A_L^{\rho_L - 1}}{A^F(A_L)}, \qquad \pi_\ell^0(\mathbf{a}) = \frac{(m_\ell - \mu_L \ln a_\ell)\, a_\ell A_L^{\rho_L - 1}}{1 + A_L^{\rho_L}},$$

below and at or above the cutoff, respectively, and, where the margin is positive, it jumps up at the cutoff, because $A^F(A_L) > 1 + A_L^{\rho_L}$. The pricing stage has an equilibrium, and every equilibrium is of one of three kinds:

(i) Accommodation. *The leaders' first-order conditions on the active branch, with markups* $\mu_L/(1-\chi_F s_{\ell|L})$, $\chi_F = 1-\rho_L+\rho_L S_L\theta_F$ *and* $\theta_F = 1/(1+\gamma S_F)$, *have exactly one solution,* $\mathbf{a}^F$, *with sub-aggregate* $A_L^F(P_L)$. *It is an equilibrium if* $A_L^F(P_L) < A_L^\dagger$ *and no leader gains by raising its action to the cutoff or beyond.*

(ii) Deterrence. *A profile with* $\sum_\ell a_\ell = A_L^\dagger$ *is an equilibrium if and only if no leader gains by raising its action,* $a_\ell \geq \widetilde{r}^0_{\ell,L}(A_L^\dagger)$, *or by lowering it and letting followers in,* $\pi^0_\ell(\mathbf{a}) \geq \sup_{0\leq b<a_\ell}\pi^F_\ell(b,\mathbf{a}_{-\ell})$. *The set* $\mathcal{D}(P_L)$ *of these profiles is compact, and on it* $N=0$ *and* $A = 1+(A_L^\dagger)^{\rho_L}$.

(iii) Ordinary exclusion. *The no-follower equilibrium* $\mathbf{a}^0$, *with sub-aggregate* $A_L^0(P_L) = A_L(0)$ *(Lemma 6(iii)), is an equilibrium if and only if* $A_L^0(P_L)\geq A_L^\dagger$, *and then it is the only one.*

Write $\mathcal{E}(P_L)$ *for the set of equilibria, and* $C(P_L)$ *for* $\{\mathbf{a}^F\}$ *if* $\mathbf{a}^F$ *is an equilibrium below the cutoff and* $\varnothing$ *otherwise.*

Proof. Accommodation. Below the cutoff the follower response is smooth near any profile, so in any equilibrium below the cutoff each leader satisfies its active-branch first-order condition. With $a_\ell = A_L s_{\ell|L}$ these read $\ln(A_L s)+1/(1-\chi_F s) = m_\ell/\mu_L$. The left side rises in s and in χ_F, and χ_F rises with A_L because S_L rises and S_F falls along the locus (Lemma 6(i)); so, as in Lemma 6(iii), exactly one profile, $\mathbf{a}^F$, solves them. By B2 no deviation that keeps followers active pays, so only deviations across the cutoff remain, which is the test in (i).

Deterrence. At the cutoff followers stay out (B6). A higher action keeps them out, and on that branch profit is single-peaked (B2), so it does not pay if and only if $a_\ell \geq \widetilde{r}^0_{\ell,L}(A_L^\dagger)$; a lower action lets them in, which is the second test. The set $\mathcal{D}(P_L)$ is bounded, since $\sum_\ell a_\ell = A_L^\dagger$, and closed, since π^F_ℓ extends continuously to the cutoff, so the supremum in the second test is a maximum over $b\in[0,a_\ell]$ and is continuous in $\mathbf{a}$.

Ordinary exclusion. Above the cutoff small deviations keep followers out, so leaders play $\mathbf{a}^0$, which needs $A_L^0(P_L) > A_L^\dagger$; a deviation $b < a^0_\ell$ that lets followers in has a margin above the positive markup at $\mathbf{a}^0$, so it earns $\pi^F_\ell(b,\mathbf{a}^0_{-\ell}) < \pi^0_\ell(b,\mathbf{a}^0_{-\ell}) \leq \pi^0_\ell(\mathbf{a}^0)$. If $A_L^0(P_L) = A_L^\dagger$, $\mathbf{a}^0$ passes both tests of (ii). Now let $A_L^0(P_L)\geq A_L^\dagger$. The ordinary shares (with χ_0) sum to the normalized reply sum of B4, which falls through one at A_L^0, so below the cutoff they sum to more than one; so do the larger shares with χ_F (followers lower S_L, and $\theta_F<1$, so $\chi_F<\chi_0$), and $\mathbf{a}^F$ is not below the cutoff. At the cutoff the ordinary replies sum to at least $A_L^\dagger$, with equality only if $A_L^\dagger = A_L^0$, so no profile on the cutoff other than $\mathbf{a}^0$ passes the first test of (ii). Hence $\mathbf{a}^0$ is the only equilibrium. If instead $A_L^0(P_L) < A_L^\dagger$, then $\mathbf{a}^0$ lies below the cutoff, where every equilibrium is $\mathbf{a}^F$; but $\mathbf{a}^0\neq\mathbf{a}^F$, since at the same A_L and shares the share equation cannot hold with both χ_0 and $\chi_F<\chi_0$. So $\mathbf{a}^0$ is an equilibrium only if $A_L^0(P_L)\geq A_L^\dagger$.

Existence. Let $\bar{a}_\ell = e^{m_\ell/\mu_L}$, $h_\ell(a) = (m_\ell - \mu_L \ln a)\, a$ with $h_\ell(0) = 0$, and $f(A_L) = A_L^{\rho_L - 1}/A(A_L)$, where $A(A_L)$ is the aggregate after followers respond, so $\pi_\ell(\mathbf{a}) = h_\ell(a_\ell) f(A_L)$. On the locus, A_F solves $H A_F^{\rho_F - 1} - A_F^{\rho_F} = 1 + A_L^{\rho_L}$, whose left side falls strictly from $+\infty$ to $-\infty$, so $A^F = H A_F^{\rho_F - 1}$ is continuous and strictly increasing in A_L. If $A_L^\dagger > 0$, then $A_F = \underline{A}_F$ at $A_L^\dagger$, so $A(A_L)$ is continuous except that it falls at the cutoff from $1 + (A_L^\dagger)^{\rho_L} + \underline{A}_F^{\rho_F}$ to $1 + (A_L^\dagger)^{\rho_L}$, the value the cutoff carries (B6). Hence f is upper semicontinuous, and $0 < f(A_L) \le A_L^{\rho_L - 1}$. On $B = \prod_\ell [0, \bar{a}_\ell]$ let $\mathcal{P}(\mathbf{a}) = f\big(\sum_\ell a_\ell\big) \prod_\ell h_\ell(a_\ell)$ and $\mathcal{P}(\mathbf{0}) = 0$. Each h_ℓ is continuous and nonnegative on $[0, \bar{a}_\ell]$, so $\mathcal{P}$ is upper semicontinuous where $A_L > 0$; and if $A_L < e^{-1}$, then $h_\ell(a_\ell) \le A_L(|m_\ell| + \mu_L |\ln A_L|)$, because $a|\ln a|$ rises on $(0, e^{-1}]$, so $\mathcal{P} \le A_L^{L + \rho_L - 1} \prod_\ell (|m_\ell| + \mu_L |\ln A_L|) \to 0$. So $\mathcal{P}$ attains its maximum on B at some $\mathbf{a}^{\mathcal{P}}$ with $\mathcal{P}(\mathbf{a}^{\mathcal{P}}) > 0$, hence $0 < a_\ell^{\mathcal{P}} < \bar{a}_\ell$. For $b \in (0, \bar{a}_\ell]$, whether or not b crosses the cutoff,

$$0 \ge \mathcal{P}(b, \mathbf{a}_{-\ell}^{\mathcal{P}}) - \mathcal{P}(\mathbf{a}^{\mathcal{P}}) = \Big[\prod_{j \ne \ell} h_j(a_j^{\mathcal{P}}) \Big] \Big[\pi_\ell(b, \mathbf{a}_{-\ell}^{\mathcal{P}}) - \pi_\ell(\mathbf{a}^{\mathcal{P}}) \Big],$$

with a positive first bracket, while $b = 0$ earns nothing and $b > \bar{a}_\ell$ gives a negative margin; so $\mathbf{a}^{\mathcal{P}}$ is an equilibrium. □

Lemma 8 (Ranking under B6)**.** *Under B5(c) and B6, the equilibrium selected by B5(b) for each nonempty prefix satisfies $\pi_k \ge \pi_\ell$ and $\pi_k/K_k \ge \pi_\ell/K_\ell$ for $k < \ell$. Thus (D1) holds.*

Proof. Fix $k < \ell$, so $L \ge 2$, $m_k > m_\ell$, and $0 < K_k \le K_\ell$. Equilibrium profits are positive, because a small enough action earns a positive profit, so it suffices to show $\pi_k \ge \pi_\ell$. At either smooth profile, $\mathbf{a}^F$ or $\mathbf{a}^0$, the common share equation orders shares by margin, hence orders profits $\mu_L S_L s/(1 - \chi_q s)$, with $\chi_q = \chi_F$ at $\mathbf{a}^F$ and $\chi_q = \chi_0$ at $\mathbf{a}^0$.

At deterrence write $h_j(a) = (m_j - \mu_L \ln a)a$, with peak $\widehat{a}_j = e^{m_j/\mu_L - 1}$, $\varphi_0 = (A_L^\dagger)^{\rho_L - 1}/(1 + (A_L^\dagger)^{\rho_L})$, $\varphi(z) = z^{\rho_L - 1}/A^F(z)$ for $0 < z \le A_L^\dagger$, with $A^F(A_L^\dagger)$ its active-side limit, $\underline{a}_j = \widetilde{r}_{j,L}^{\,0}(A_L^\dagger)$, and

$$\Psi_j(a) = \max_{0 \le b \le a} h_j(b) \varphi(A_L^\dagger - a + b), \qquad T_j(a): \quad \varphi_0 h_j(a) \ge \Psi_j(a).$$

On $\mathcal{D}(P_L)$, $\pi_j = \varphi_0 h_j(a_j)$ and a deviation $b < a_j$ earns $h_j(b)\varphi(A_L^\dagger - a_j + b)$, so by Lemma 7(ii) each firm's cutoff feasibility depends only on its own action: $a_j \ge \underline{a}_j$ and $T_j(a_j)$. Since there are at least two firms, $0 < a_j < A_L^\dagger$; φ is continuous and strictly decreasing, since $z^{\rho_L - 1}$ is nonincreasing and A^F is strictly increasing (Lemma 6(i)), and $\varphi(A_L^\dagger) < \varphi_0$. Four facts supply the feasible reallocations used below.

Downward closure. The test forces a positive margin. With $M = m_j - \mu_L \ln a$,

$$(\ln h_j)'' = -\frac{M^2 - \mu_L M + \mu_L^2}{a^2 M^2} < 0.$$

For every fixed reduction $\delta > 0$, the ratio $h_j(a-\delta)/h_j(a)$ increases with a. In the test, the multiplier $\varphi(A_L^\dagger - \delta)/\varphi_0$ is independent of a. Reducing an admissible action retains every remaining deviation inequality, so T_j is downward closed.

Margin ordering. $T_\ell(a)$ implies $T_k(a)$. Raising the margin adds to each deviation inequality

$$(m_k - m_\ell)[a\varphi_0 - b\varphi(A_L^\dagger - a + b)] > 0.$$

Indeed $b\varphi(z) = (b/z)S_L(z)$ with $z = A_L^\dagger - a + b$, where $S_L(z) = z^{\rho_L}/A^F(z)$ rises along the locus (Lemma 6(i)), so $b\varphi(z)$ increases up to $a\varphi(A_L^\dagger) < a\varphi_0$ as b approaches a.

Reply bounds. $\underline{a}_j = A_L^\dagger s_j$, where s_j solves (46) at $T = A_L^\dagger$ and $N = 0$, with the common χ_0. A higher margin gives a higher s_j, so $\underline{a}_\ell < \underline{a}_k$; and the ordinary markup $m_k - \mu_L \ln \underline{a}_k = \mu_L/(1-\chi_0 s_k)$ exceeds μ_L, so $\underline{a}_k < e^{m_k/\mu_L - 1} = \widehat{a}_k$.

Deviation values. Ψ_j is continuous and nondecreasing on $(0, A_L^\dagger)$: increasing a retains old positive-payoff deviations and lowers their total $A_L^\dagger - a + b$, increasing φ. Also $\Psi_k(a) \geq \Psi_\ell(a)$ by the margin order.

Priority prohibits raising h_k while keeping earlier firms' actions fixed. If $a_k < a_\ell$, exchanging the two actions is feasible: downward closure preserves ℓ's test, margin ordering preserves k's, and both lower bounds hold. Hence $h_k(a_k) \geq h_k(a_\ell) > h_\ell(a_\ell)$. If $a_\ell \leq a_k \leq \widehat{a}_k$, the same profit inequality follows from the increase of h_k up to its peak. Otherwise $a_\ell \leq a_k$ and $a_k > \widehat{a}_k$. Then T_ℓ must bind: if slack, a small transfer of action from k to ℓ retains both tests and lower bounds while raising h_k, contradicting priority. Consequently

$$\varphi_0 h_k(a_k) \geq \Psi_k(a_k) \geq \Psi_k(a_\ell) \geq \Psi_\ell(a_\ell) = \varphi_0 h_\ell(a_\ell).$$

Operating-profit ranking and ordered-cost ratio ranking follow. A singleton has no pairwise ranking to check. □

Proposition 3L. ***Nested Multiple-Leader Equilibrium and Global Uniqueness.*** *Under the nested primitives and assumptions B2, B5, and B6, the legal-entry, legal-price-commitment, and illegal-entry game has a unique selected subgame-perfect equilibrium. Deterrence priority B5(b) applies to the whole pricing equilibrium set $\mathcal{E}(P_L)$ of each prefix. The selected continuation satisfies (D1), and its legal set is*

$$L^* = \max\big(\{0\} \cup \{L \in \{1, \ldots, \bar{L}\} : \Gamma_L^N > 1\}\big), \qquad \Gamma_L^N \leq 1 \text{ for } L > L^*.$$

Every entrant's ratio in the final set exceeds one. The continuation can be accommodation, ordinary exclusion, or deterrence at the cutoff. Without B6 there is a unique equilibrium, the fringe is always active, and deterrence priority is unnecessary.

Proof. For a nonempty prefix, Lemma 7 gives $\mathcal{E}(P_L) = \{\mathbf{a}^0\}$ if $A_L^0(P_L) \geq A_L^\dagger$ and $\mathcal{E}(P_L) = C(P_L) \cup \mathcal{D}(P_L)$ otherwise; the empty prefix has the follower continuation of Lemma 6(i)–(ii).

Selection. By the downward-closure and deviation-value steps in the proof of Lemma 8, each firm's cutoff test holds on an interval $[0, U_j]$, so

$$\mathcal{D}(P_L) = \left\{\mathbf{a} : \ \textstyle\sum_j a_j = A_L^\dagger,\ \widetilde{r}_{j,L}^{0}(A_L^\dagger) \leq a_j \leq U_j\right\}$$

is compact and convex. On it $\pi_j = \varphi_0 h_j(a_j)$, with φ_0 and h_j as in that proof and $h_j''(a) = -\mu_L/a < 0$, so lexicographic maximization fixes a unique a_1, then a_2, and so on. Comparing this profile with $\mathbf{a}^F$ (when $C(P_L) \neq \varnothing$) by the profit vector, with ties going to deterrence, selects one profile, so the lower-price tie rule of B5(b) is never used. Lemma 8 gives (D1) at that profile, and Lemma 2 selects P_{L^*}. Followers play a_f^0 at price $c_F + \mu_F$, and $A^* = 1 + (A_L^*)^{\rho_L} + (A_F^*)^{\rho_F}$.

Case 1: Legal firms and illegal sellers enter ($L^* > 0$, selected profile $\mathbf{a}^F$, $A_L^F(P_{L^*}) < A_L^\dagger$): $A_L^* = A_L^F(P_{L^*})$, and $A_F^* > \underline{A}_F$ solves

$$\left(1 + (A_L^*)^{\rho_L} + A_F^{\rho_F}\right) A_F^{1-\rho_F} = H, \qquad N^* = A_F^*/a_f^0. \tag{48}$$

Leaders commit before entry and expect followers to move along the locus, so a price cut raises A by only the fraction $\theta_F = 1/(1 + \gamma S_F)$ of its direct effect. They price with $\chi_F = 1 - \rho_L + \rho_L S_L \theta_F$ (Lemma 7(i)), below the simultaneous χ, so at given shares legal markups are lower than in Proposition 3S. Given A_L^*, the follower mass is the single crossing in Lemma 6(i)–(ii), and B5(b) selects $\mathbf{a}^F$ only if $\mathcal{D}(P_{L^*})$ is empty or $\mathbf{a}^F$'s profit vector beats every deterrence profile.

Case 2a: Legal firms enter and illegal sellers do not (ordinary exclusion) ($L^* > 0$, $A_L^0(P_{L^*}) \geq A_L^\dagger$): the ordinary profile is the only equilibrium, and

$$A_L^* = A_L^0(P_{L^*}), \qquad N^* = 0, \qquad A^* = 1 + (A_L^*)^{\rho_L}. \tag{49}$$

Ordinary prices already put the legal total at or above the cutoff, so a minimum mass cannot profit even without commitment; Lemma 7(iii) gives $\mathcal{E}(P_{L^*}) = \{\mathbf{a}^0\}$, and no selection is needed. The outcome is that of the first part of Case 2 in Proposition 3S, with the aggregate set by legal primitives.

Case 2b: Legal firms deter and illegal sellers do not enter ($L^* > 0$, $A_L^0(P_{L^*}) < A_L^\dagger$, selected profile in $\mathcal{D}(P_{L^*})$):

$$A_L^* = A_L^\dagger, \qquad N^* = 0, \qquad A^* = 1 + (A_L^\dagger)^{\rho_L}. \tag{50}$$

Leaders raise their total from $A_L^0(P_{L^*})$ to the cutoff $A_L^\dagger$, where a minimum mass would exactly break even and, by B6, stays out; the aggregate $1 + (A_L^\dagger)^{\rho_L}$ exceeds the ordinary $1 + A_L^0(P_{L^*})^{\rho_L}$, so the threat of entry raises consumer surplus. This is the nested counterpart of limit pricing in Proposition 1L. Every profile in $\mathcal{D}(P_{L^*})$ gives the same A^*, N^*, and consumer surplus but divides

profit differently and can coexist with accommodation, which is why B5(b) is needed; without B6 there is no cutoff and this case is empty.

Case 3: Illegal sellers enter and legal firms do not ($L^* = 0$, $A_L^\dagger > 0$): $A_F^* > \underline{A}_F$ solves

$$\left(1 + A_F^{\rho_F}\right) A_F^{1-\rho_F} = H. \tag{51}$$

With no legal firm there are no prices to commit to, so the outcome is that of Case 3 of Proposition 3S, with $N^* = A_F^*/a_f^0$ and $A_L^\dagger > 0$ the activation test of Lemma 6(ii) at $A_L = 0$. As there, no candidate enters because $\Gamma_L^N \leq 1$ for every L, here with leadership profits.

Case 4: Neither enters ($L^* = 0$, $A_L^\dagger = 0$): $A_L^* = A_F^* = N^* = 0$ and $A^* = 1$. As in Case 4 of Proposition 3S, a minimum mass loses even against the outside good alone, and $\Gamma_L^N \leq 1$ for every L.

Without B6. Followers are active at every legal total, because $\widehat{\pi}_F \to \infty$ as $A_F \to 0$, so leader payoffs have no kink. The share-sum argument in the proof of Lemma 7(i), now on the whole locus, gives the unique profile $\mathbf{a}^F$, a global best response by B2; the common share equation orders it by margin, which is (D1), and only Cases 1 and 3 remain, with no selection. □

Remark 4 (Comparing the two timings without B6). For each legal set both timings have a unique continuation; queue uniqueness does not require B5(c), although the stated maximum-prefix formula uses the within-prefix ranking it supplies. A leader expects follower entry to offset part of a price cut, so $\chi_F < \chi$ and its markup is lower at given shares. This is a conduct comparison at given shares, not a universal ordering of prices or pass-through across different equilibria. Free entry pins a locus rather than a fixed aggregate. As $\rho_L, \rho_F \to 1$ along an active-fringe path with a positive limiting illegal share, $\gamma S_F \to \infty$ and $\chi_F \to 0$, so the leader markup tends to μ, the logit benchmark with full crowd-out. Vanishing-fringe paths need a separate limiting analysis.

A.2.2 Other Demand Systems

The results in Sections 3.1–3.5 use logit demand only through Lemma 1. For a general aggregative game with payoffs $\pi_j(a_j, A)$ and aggregate $A = A_0 + \sum_\ell a_\ell + N a_f$, where A_0 is the outside good's action, the corresponding conditions are A1–A4 of Section 3.3, which logit satisfies (Lemma 1; A1 at every action with a positive markup, which includes every best reply).

Remark 5 (Free entry pins the aggregate). Suppose A1 and A2 hold, illegal sellers are identical and atomistic, and an illegal seller's maximized operating profit $\widehat{\pi}_F(A)$ is continuous and strictly decreasing in A, with $\widehat{\pi}_F(\bar{A}) = K_F$. Then $A = \bar{A}$ in every equilibrium with $N > 0$: one seller moves neither A nor any price, so active sellers and entrants all earn $\widehat{\pi}_F(A) - K_F$, which free entry sets to zero, and $\bar{A}$ is the only root. If, in addition, each legal firm's reply at $\bar{A}$ is strictly decreasing

in its tax, then while illegal sellers remain active a higher legal tax leaves A and the illegal price unchanged and, for a given nonempty legal set, raises N: the illegal action maximizes $\pi_f(\cdot, \bar{A})$, which contains no legal primitive, so a_f and the illegal price are unchanged, while legal actions fall and $Na_f = \bar{A} - A_0 - \sum_\ell a_\ell$ rises.

CES Bertrand. With elasticity of substitution $\eta > 1$, taste shifter v_j, marginal cost c_j and expenditure Y, firm j's action is $a_j = v_j p_j^{1-\eta}$, the aggregate is $A = A_0 + \sum_\ell a_\ell + Na_f$, and $\pi_j = Y\big(1 - (c_j + \tau)/p_j\big)a_j/A$, with the tax τ on legal firms only. A legal firm's Nash price is $(c_\ell + \tau)\,\varepsilon/(\varepsilon - 1)$, where $\varepsilon = \eta - (\eta - 1)s$ and $s = a_\ell/A$ is its share. An illegal seller's price is $c_F\,\eta/(\eta - 1)$, and free entry pins $\bar{A} = Y\widehat{r}_f/(\eta K_F)$, where $\widehat{r}_f = v_F\big(c_F\eta/(\eta - 1)\big)^{1-\eta}$ is an illegal seller's action.

Differentiated linear Cournot. With inverse demand $p_j = v_j - b\,q_j - d\,Q$, aggregate $A = Q = \sum_\ell q_\ell + Nq_f$ and legal margin $m_\ell = v_\ell - c_\ell - \tau$, a legal firm produces $q_\ell = (m_\ell - dQ)/(2b + d)$ (Nash) or $(m_\ell - dQ)/(2b)$ (leader), an illegal seller produces $q_f = (v_F - c_F - dQ)/(2b)$, and free entry pins $\bar{Q} = (v_F - c_F - 2\sqrt{bK_F})/d$.

CES satisfies A1–A4. Differentiated linear Cournot satisfies A1, A2 (for $b > 0$) and A4 but not A3: a firm's output falls as total output rises. Candidates are ranked by quality-adjusted cost under CES and by $v - c$ under Cournot, in place of A5(c). Uniqueness for CES and differentiated Cournot has been checked only numerically, and in both A rises with N and with each added legal firm. A Cournot candidate with $m_\ell \le d\bar{Q}$ cannot profit while illegal sellers are active, so it does not enter.

In all three systems free entry pins the aggregate while illegal sellers are active (Remark 5), so a legal tax leaves it unchanged, changes only the composition of sales, and raises illegal entry. Total units consumed stay fixed under logit (they equal $1 - 1/\bar{A}$) and under linear Cournot (they equal Q), but can move either way under CES. Consumer surplus stays fixed under logit and CES but changes with composition under Cournot, where $CS = \frac{b}{2}\sum_j q_j^2 + \frac{d}{2}Q^2$, so more, smaller illegal sellers change surplus at fixed Q. Legal pass-through in the simultaneous game is below one under logit (Lemma 4(iii)) and equals $b/(2b + d)$ under Cournot; under CES it is $g/[1 + (\eta - 1)^2 s/(\varepsilon(\varepsilon - 1))]$, with $s = a_\ell/A$, $\varepsilon = \eta - (\eta - 1)s$ and $g = \varepsilon/(\varepsilon - 1)$, which can exceed one. In the leader game it is 1 under logit, $\eta/(\eta - 1)$ under CES and $1/2$ under Cournot. Homogeneous Cournot ($b = 0$) needs rising marginal cost for illegal sellers; with the aggregate pinned, a leader's profit is linear in output, so leaders always push illegal sellers out.

A.3 Empirical Model: Methods and Definitions

A.3.1 The Nine Models

Table 2 in Section 4.4 lists the nine models. Each is estimated with MNL demand ($\sigma = 0$) and with nested logit demand; nested logit is the preferred demand specification. Models 1–3 never

estimate N_{st}. Models 4–6 obtain it by substituting the zero-profit condition (Appendix A.3.2). Models 7–9 recover it from several readings at every candidate parameter vector (Appendix A.3.3) and differ in how many readings enter (two in Model 7, four in Models 8 and 9) and how they are weighted (Appendix A.3.4).

Models 2–9 use the converted tax wedge (Appendix A.3.6). Demand, supply and each arrest row have their own constant, region effects and year effects; the zero-profit row has none. Weighting matrices and standard errors use heteroskedasticity-robust covariances of observation-level moment contributions. No lower bound is imposed on N_{st}, σ is restricted to $[0, 0.999]$, and loadings are unrestricted in sign.

A.3.2 Substitution Route, Models 4–6

Benchmarks. Demand only (Model 1) drops $(1-\sigma)\ln N_{st}$ from illegal demand, so the illegal intercept ϕ and the fixed effects absorb average entry. No entry (Model 2) adds the supply rows, neither of which involves N_{st}. Legal only (Model 3) drops the illegal rows, so illegal buyers are indistinguishable from non-buyers and the outside share becomes $1 - S_{Lst}$; substitution toward illegal sellers then counts as substitution toward not buying.

Substituting zero profit. The entry cost is $\ln K_{Fst} = \ln K_0 + \widetilde{W}'_{Fst}\theta + \zeta_{Fst}$, with the eight cost shifters $\widetilde{W}_{Fst}$ measured from their means; Model 4 sets $\theta = 0$. Zero profit gives $\ln N_{st} = z_{st} - \zeta_{Fst}$ with $z_{st} = \ln\frac{1-\sigma}{\alpha} + \ln M_{st} + \ln S_{Fst} - \ln K_0 - \widetilde{W}'_{Fst}\theta$, and substitution sets $\ln N_{st} = z_{st}$ in illegal demand and the arrest rows:

$$\ln\frac{S_{Fst}}{s_{0st}} = (1-\sigma)z_{st} + x'_{Fst}\beta - \alpha p_{Fst} + \phi + f^D_{rt} + \left[\xi_{Fst} - (1-\sigma)\zeta_{Fst}\right], \tag{52}$$

$$y_{gst} = \lambda_g z_{st} + f^g_{rt} + \left[\iota_{gst} - \lambda_g\zeta_{Fst}\right], \qquad g = 1, 2. \tag{53}$$

The bracketed terms are composite errors. Legal demand and both supply rows are unchanged, because none of them involves N_{st}. The illegal-demand instruments must be uncorrelated with $\xi_{Fst} - (1-\sigma)\zeta_{Fst}$, and the arrest instruments with $\iota_{gst} - \lambda_g\zeta_{Fst}$; uncorrelatedness with each component shock is a simple sufficient condition (Bollen, 1996). An instrument valid for the demand shock alone fails if it is also correlated with the part of entry cost that the shifters do not capture. There is no separate zero-profit block: its residual is zero by construction, and its shock has moved into the other rows. Substitution is the recovery that puts all the weight on zero profit: $a_Z = 1$ and $a_D = a_1 = a_2 = 0$ in the notation of Section 4.3.

K_0 is fixed, not estimated. Raising it by a factor e^c lowers every $\ln N_{st}$ by c, shifts ϕ by $(1-\sigma)c$ and each arrest constant by $\lambda_g c$, and leaves every residual and every other estimate unchanged. Absolute entry is therefore conditional on K_0, while ratios of N across markets and years and the

percentage responses of N do not depend on it. The same holds on the concentration route with fixed or frozen weights.

In Models 2 and 4–9 the cost equation has an intercept shift for illegal sellers,

$$\ln c_{kst} = w'_{kst}\gamma + \gamma_F \, \mathbb{1}\{k = F\} + f^S_{rt} + \eta_{kst}, \qquad k = j, F. \tag{54}$$

Illegal sellers pay no license, testing, traceability or packaging costs, so they can face a different cost level from licensed retailers with the same shifters. Without γ_F, only the illegal markup $(1-\sigma)/\alpha$ can close the gap between illegal and legal costs, and the fit pushes σ toward its upper bound of 0.999. One constant cannot track the gap between statutory and converted rates, which widens over 2014–2016 (Table 17), so γ_F may partly absorb the conversion rather than measure a cost advantage.

Model 6 adds the two arrest rows (53) to Model 5, with separate loadings and no direct controls; N_{st} is still the zero-profit value. Given the other parameters, the cost coefficients and loadings enter the moments linearly and are obtained by projection at each trial point (variable projection; Golub and Pereyra, 1973), and the remaining parameters are searched within their bounds.

A.3.3 Concentration Route, Models 7–9

Readings and recovery. Write $n_{st} \equiv \ln N_{st}$. Equation (19) gives four readings, $d_{st} = (1-\sigma)n_{st} + \xi_{Fst}$, $z_{st} = n_{st} + \zeta_{Fst}$ and $y^*_{gst} = \lambda_g n_{st} + \iota_{gst}$. Stack the readings in $r_{st} = (d_{st}, z_{st}, y^*_{1st}, y^*_{2st})'$, their loadings on entry in $b = (1-\sigma, 1, \lambda_1, \lambda_2)'$ and the shocks in $e_{st} = (\xi_{Fst}, \zeta_{Fst}, \iota_{1st}, \iota_{2st})'$, so the four equations read $r_{st} = b\, n_{st} + e_{st}$; Model 7 uses the first two rows. Fixed effects and, in Models 8 and 9, the indicator controls are removed from both the arrests and the entry they are compared with, so the indicator residual is $M_{D_g}(y_g - \lambda_g n)$, where M_{D_g} removes the fixed effects and controls of row g. Let P be a positive-semidefinite 4×4 weighting matrix with $b'Pb > 0$. In Models 7 and 8 it is diagonal, $P = \mathrm{diag}(\omega_D, \omega_Z, \omega_1, \omega_2)$, with $\omega_1 = \omega_2 = 0$ in Model 7; in Model 9 it is full, the inverse covariance of the readings, estimated once from a first-step fit and then held fixed. Given Θ, recovery solves

$$\widehat{n}(\Theta) = \arg\min_n F(\Theta, n), \qquad F(\Theta, n) = \tfrac{1}{2}\textstyle\sum_{st}(r_{st} - b\, n_{st})'P(r_{st} - b\, n_{st}), \tag{55}$$

jointly with the nuisance coefficients. With diagonal P, each market's term in F is half the weighted squared gaps (20).

Recovery with cross weights. Write $\tilde{e}_{st}(n) = r_{st} - b\,n$ for the gaps at a candidate n. The criterion in (55) judges both their sizes and how they move together:

$$\tilde{e}'_{st} P\, \tilde{e}_{st} = \sum_i P_{ii}\, \tilde{e}^2_{i,st} + 2 \sum_{i<j} P_{ij}\, \tilde{e}_{i,st}\, \tilde{e}_{j,st}. \tag{56}$$

Market by market the solution and its error are

$$\widehat{n}_{st} = \frac{b' P\, r_{st}}{b' P\, b} = a' r_{st}, \qquad \widehat{n}_{st} - n_{st} = a' e_{st}, \qquad a = \frac{Pb}{b' P b}, \tag{57}$$

where $a = (a_D, a_Z, a_1, a_2)'$ holds the weight on each reading and $a'b = 1$, so entry cancels. With diagonal P the cross terms vanish, the weights are $a_h = \omega_h b_h / \Delta$ as in Section 4.3, with $\Delta = \sum_k \omega_k b_k^2$, and $\widehat{n}_{st}$ is the weighted average (21), the weighted least-squares factor score of Bartlett (1937) when the ω_h are inverse error variances. With a full P the cross weights P_{ij} enter. If P inverts the covariance of the reading errors, a positive covariance, such as an enforcement shock shared by the two arrest series, produces a negative cross weight. One series then partly nets the shared shock out of the other, and a weight a_j can be negative even when every loading is positive. In every case the residuals entering the moments are the combined errors

$$\widehat{u}_{st} = r_{st} - b\,\widehat{n}_{st} = (I - b\,a')\, e_{st}, \qquad \text{so} \qquad \widehat{u}_{k,st} = e_{k,st} - b_k\, a' e_{st}, \tag{58}$$

which is Equation (23) with $\bar{e}_{st} = a' e_{st}$. With diagonal weights, for example, the recovered illegal-demand residual is $(1 - (1-\sigma)a_D)\xi_{Fst} - (1-\sigma)(a_Z \zeta_{Fst} + a_1 \iota_{1st} + a_2 \iota_{2st})$.

Combined shocks under covariance weights. With a full P (Model 9), each weight also depends on the cross weights. For the sales-arrest reading, for example,

$$a_1 = \frac{P_{1D}(1-\sigma) + P_{1Z} + P_{11}\lambda_1 + P_{12}\lambda_2}{b' P b}, \tag{59}$$

which is $\omega_1 \lambda_1 / \Delta$, with $\Delta = b'Pb$, when the cross weights are zero. The moments (34) are built on the same combined shocks, $\xi_{Fst} - (1-\sigma)\bar{e}_{st}$, $\zeta_{Fst} - \bar{e}_{st}$ and $\iota_{gst} - \lambda_g \bar{e}_{st}$, with $\bar{e}_{st} = a_D \xi_{Fst} + a_Z \zeta_{Fst} + a_1 \iota_{1st} + a_2 \iota_{2st}$; what changes is how much of each shock enters $\bar{e}_{st}$. Suppose an enforcement campaign adds a common shock χ_{st} to both arrest series, $\iota_{gst} = \chi_{st} + \iota^0_{gst}$, so that χ_{st} enters $\bar{e}_{st}$ with weight $a_1 + a_2$. Diagonal weights treat the two series as independent readings and so count the campaign twice. Take a P that inverts a covariance in which only the two arrest series move together. Then $P_{12} < 0$, which discounts what the second series adds. When the two series are symmetric, $a_1 + a_2$ falls, so less of the campaign passes into recovered entry and, through it, into the illegal-demand and zero-profit residuals. When the shared shock is large relative to

the possession series' own noise and possession arrests load more heavily on entry ($\lambda_2 > \lambda_1$), the sales-arrest weight a_1 can turn negative: the sales series is then used to remove the shared shock from the possession series rather than to measure entry. The campaign stays in the arrest residuals themselves, so the arrest instruments must still be uncorrelated with enforcement.

The validity condition is the one in Section 4.3: each equation's instruments must be uncorrelated with every shock that has a nonzero weight in its combined shock. A negative weight flips the sign of a component but does not relax the requirement that it be valid, and the estimates do not rely on biases in two components offsetting. Because P is frozen, the weights move with Θ only through the loadings $(1-\sigma, 1, \lambda_1, \lambda_2)$, so the response of recovered entry has no weight-derivative term. When the weights are computed from the readings (Model 7 at every candidate, Model 9 once at its first-step fit), a depends on the data and the argument applies to its population limit. A covariance calculation cannot tell shared entry from shared measurement or enforcement noise.

In the MNL example of Section 4.3, the demand residual $(\xi - \zeta)/2$ carries entry-cost noise, so for a demand instrument Z, $E[Z\widehat{u}_D] = \frac{1}{2}E[Z\xi] - \frac{1}{2}E[Z\zeta]$.

Recovery minimizes Equation (55), not the GMM criterion, so in general $\widehat{n}(\Theta)$ is not the n that minimizes $Q(\Theta, n)$, the GMM criterion with entry held at n. The estimator is GMM with a recovered latent state, not the GMM criterion minimized over entry. Its justification is that the recovered residuals satisfy valid moments that identify Θ. Recovering an object inside the objective at every candidate is standard (Rust, 1987; Berry et al., 1995; Nevo, 2000), but there the inner problem is solved exactly, using a contraction mapping.

The first-order condition of recovery, $b'P\widehat{u}_{st} = 0$, holds in every market. For an instrument Z that enters the four reading blocks after the same projection, summing over markets gives

$$\textstyle\sum_j (Pb)_j \, g_{j,Z} = 0, \tag{60}$$

where $g_{j,Z}$ is the moment of block j with instrument Z; with diagonal P the coefficients are $\omega_j b_j$. A fixed combination of four moments (two in Model 7) is then identically zero and carries no information. Models 7 and 8 have nine such identities, one for the constant and one for each entry-cost shifter; under Model 9's full P they hold only approximately. Left in, they have near-zero variance and so receive very large weight. The estimator recomputes entry at every Θ, and it makes three changes:

(i) it projects the moments onto an orthonormal basis of the moment coordinates left after the nine identities are removed;

(ii) it corrects each observation's moment contribution for the fitted entry and nuisance coefficients (Newey and McFadden, 1994; Murphy and Topel, 1985);

(iii) it rebuilds the step-two weight on the remaining coordinates.

A.3.4 Weighting the Readings, Models 7–9

Three covariance objects enter the estimator:

- the recovery weighting P, which fixes a and so the mixture in (58);
- the covariance of the moments, whose inverse is the weight W_G and which also gives the standard errors;
- the covariance Σ of the four shocks.

The models differ only in P. Model 7 uses the first two readings, with diagonal weights ω_D and ω_Z recomputed at every candidate Θ as the inverse of what remains of each reading after the common movement is attributed to entry. Model 8 uses diagonal weights on all four readings, calibrated once and held fixed. Model 9 uses a full P, the inverse covariance of the four readings, estimated once from a first-step fit and then held fixed. A working weighting matrix P can combine correlated readings without identifying Σ (Liang and Zeger, 1986). If entry is uncorrelated with the shocks, $\mathrm{Var}(r_{st}) = v\, bb' + \Sigma$, where $v = \mathrm{Var}(n_{st})$, and $(v + c,\ \Sigma - c\, bb')$ gives the same covariance, so the readings alone cannot separate entry variance from shock covariance.

A.3.5 Computation: Profiled Objective

Models 7–9 minimize $Q_P(\Theta)$ in Equation (24), with $\widehat{n}(\Theta)$ in place of $\widehat{\ln N}(\Theta)$, recomputing the readings, any Θ-dependent weights and $\widehat{n}(\Theta)$ at every evaluation. An interior estimate solves

$$\nabla Q_P(\Theta) = Q_\Theta + \widehat{n}'_\Theta\, Q_n = 0, \tag{61}$$

where Q_Θ and Q_n are the partial derivatives of $Q(\Theta, n)$ at $n = \widehat{n}(\Theta)$, and $\widehat{n}_\Theta = \partial\widehat{n}/\partial\Theta'$ is the response of recovered entry. The second term does not vanish, because $\widehat{n}$ sets the derivative F_n of the recovery criterion to zero, not Q_n.

Differentiating the first-order condition of recovery, $F_n\big(\Theta, \widehat{n}(\Theta)\big) = 0$, gives the response $\widehat{n}_\Theta = -F_{nn}^{-1} F_{n\Theta}$. A positive-definite F_{nn} guarantees a unique entry solution, but it does not guarantee that the moments identify Θ.

A.3.6 Taxes

Before July 1, 2015, the retail 25% excise was the retailer's own tax on its receipts, which were also the sales-tax base. From July 1, 2015, a single 37% excise is imposed on the buyer, added to the selling price and excluded from the sales-tax base. The sales-tax rate t is the state rate of 6.5% plus the local rate. Upstream excise is already in observed wholesale prices. Take \$100 of selling price and $t = 9\%$. In 2014 the customer pays \$109 and the retailer keeps \$75; in 2016 the customer pays \$146 and the retailer keeps \$100. In general, the share of the customer's payment that the

retailer keeps is

$$1-\tau^{\text{pre}}=\frac{0.75}{1+t},\qquad 1-\tau^{\text{post}}=\frac{1}{1.37+t}, \tag{62}$$

which equal 0.688 and 0.685 at $t = 9\%$. The reform left the wedge almost unchanged.

The converted rate x_c is the excise as a share of the full customer payment. For a statutory rate x,

$$x_c^{2014}=\frac{x}{1+t},\qquad x_c^{2016}=\frac{x}{1+x+t}. \tag{63}$$

The state's data report retail sales and excise of $R_0 = \$138{,}979{,}645$ and $E_0 = \$34{,}744{,}912$ for January–June, and $R_1 = \$218{,}595{,}740$ and $E_1 = \$80{,}880{,}424$ for July–December. The annual rate is the ratio of excise to reported sales, $x_0 = (E_0+E_1)/(R_0+R_1) = 0.323359$, not a calendar average. A counterfactual annual rate x scales both regime rates by $k = x/x_0$, keeping their ratio at 37/25, so the annual rate is a single scale factor rather than the rate charged in either half. Customer spending is $(1+t)R_0$ before July, when the excise was inside reported receipts, and $(1+t)R_1 + kE_1$ after; the excise is added to the customer's payment only after the change. Hence

$$x_c^{2015}=\frac{k(E_0+E_1)}{(1+t)(R_0+R_1)+kE_1}=\frac{x}{1+t+\frac{E_1}{E_0+E_1}x},\qquad \frac{E_1}{E_0+E_1}=0.6995. \tag{64}$$

Table 17: Statutory and converted excise rates at a combined sales-tax rate of 9%. The converted rate x_c is the excise as a share of the full customer payment (Equations (63) and (64)). For 2015, the annual rate x scales the January–June and July–December statutory rates by the same factor $k = x/0.323359$. Bold entries are the converted rates at the statutory rates in force: 25% in 2014, the annual rate of 32.34% in 2015 and 37% in 2016.

	2014	2015			2016
Statutory or annual rate x	x_c	Jan–Jun rate	Jul–Dec rate	x_c	x_c
0%	0.00%	0.00%	0.00%	0.00%	0.00%
10%	9.17%	7.73%	11.44%	8.62%	8.40%
20%	18.35%	15.46%	22.88%	16.26%	15.50%
25%	**22.94%**	19.33%	28.61%	19.76%	18.66%
30%	27.52%	23.19%	34.33%	23.08%	21.58%
32.34%	29.67%	25.00%	37.00%	**24.57%**	22.88%
37%	33.94%	28.61%	42.34%	27.43%	**25.34%**
40%	36.70%	30.93%	45.77%	29.20%	26.85%
50%	45.87%	38.66%	57.21%	34.73%	31.45%
60%	55.05%	46.39%	68.65%	39.74%	35.50%

The tax treatment sets τ_{st} in estimation and counterfactuals. At given parameters it changes recovered legal marginal cost, $c_{jst} = (1-\tau_{st})(p_{jst}-\mu_{jst})$, so estimates that use the converted wedge differ from estimates that use the statutory rates.

A.3.7 Elasticities, Diversion and Leakage: Definitions

In a county-year, store j has share $s_j = Q_j/M$; the legal and illegal sectors have $S_L = \sum_j s_j$ and $S_F = Q_F/M$; the outside good has $s_0 = 1 - S_L - S_F$; and $s_{j|L} = s_j/S_L$. Let $q = 1 - \sigma$. Prices are customer prices, $\bar{p}_L$ and $\bar{p}_C$ are quantity-weighted legal and cannabis prices, and $1 - \tau_{st}$ is the share of the customer's payment that the retailer keeps. The measures below hold other prices, parameters and demand shocks fixed, and, in Models 4–9, the number of illegal sellers N. The legal-price elasticities in Section 4.6 instead let N adjust through zero profit, $N/N_b = Q_F/Q_{F,b}$, and the response of total cannabis quantity is the response of legal plus illegal quantity.

For a legal store and an illegal seller,

$$\varepsilon_{jj} = -\frac{\alpha p_j}{q}\left[1 - \sigma s_{j|L} - q\, s_j\right], \qquad \varepsilon_{ff} = -\frac{\alpha p_F}{q}, \tag{65}$$

the second implied by the illegal pricing rule $p_F = c_F + q/\alpha$, which treats each seller as small. For a finite number of identical sellers the illegal elasticity is $-\frac{\alpha p_F}{q}[1 - (\sigma + qS_F)/N]$, which I report only where $N \geq 1$.

For a common percentage price increase in the named sector, where P_L and P_F are the legal and illegal price levels,

$$\begin{gathered}\varepsilon_L = -\alpha \bar{p}_L(1 - S_L), \qquad \varepsilon_I = -\alpha p_F(1 - S_F), \qquad \varepsilon_C = -\alpha \bar{p}_C\, s_0, \\ \frac{\partial \ln Q_F}{\partial \ln P_L} = \alpha \bar{p}_L S_L, \qquad \frac{\partial \ln Q_L}{\partial \ln P_F} = \alpha p_F S_F.\end{gathered} \tag{66}$$

When all legal prices rise, the share of lost legal quantity that goes to illegal sellers, the leakage share, is

$$D_{L\to F} = \frac{S_F}{1 - S_L} \text{ with } N \text{ fixed}, \qquad D^{\text{free}}_{L\to F} = \frac{S_F}{S_F + \sigma s_0} \text{ when illegal sellers enter},$$

and the second equals one under MNL. When one store raises its price, its lost quantity goes to other legal stores, illegal sellers and the outside good in the shares $(S_L - s_j)(q + \sigma/S_L)/B_j$, qS_F/B_j and qs_0/B_j, with $B_j = 1 - \sigma s_{j|L} - qs_j$. These shares sum to one.

The legal pricing condition gives the markup $\mu_j = q/(\alpha B_j)$ and marginal cost $c_j = (1 - \tau_{st})(p_j - \mu_j)$. The margin on retained revenue is $100\mu_j/p_j = -100/\varepsilon_{jj}$, store by store.

The tax and cost experiments change the excise rate or a cost and let legal prices re-solve the pricing condition, with N adjusting through zero profit, $N/N_b = Q_F/Q_{F,b}$, where the subscript b marks the baseline. The illegal price moves only with illegal cost. In the experiments entry responds only through zero profit. Tax leakage is the quantity illegal sellers gain after an excise increase divided by the legal quantity lost.

Annual measures sum derivatives and quantities across counties, and diversion sums gains over losses; store percentiles describe dispersion across stores, not uncertainty. Model 1 has no pricing condition, so it has no margins or equilibrium responses. Model 3's outside good includes illegal purchases, so it has no illegal-sector, cannabis, cross-price or leakage measures.

Tables 18–20 show how the estimates and responses change under alternative specifications and Table 3 gives ranges across the nine models. All numerical outcomes refer to 2016. The estimation sample is either 2014–2016 (full sample) or 2015–2016. County nesting, the baseline, puts all legal stores of a county-year in one nest. Population in demand adds log county-year population only to demand. City nesting splits legal products into parallel city nests within each county-year, alongside one illegal nest, with one freely estimated σ common to all nests. City nesting is estimated only with nested logit, because MNL has no nests.

Table 18: Estimates under alternative specifications, 2016. α is the price coefficient, σ the nesting parameter ($\sigma = 0$ under MNL), the wholesale coefficient is the coefficient on the wholesale price in the cost equation, and N is the number of illegal sellers in Washington in 2016, summed over counties.

Specification	Demand model	α	σ	Wholesale coefficient	2016 entry N
County nests; full sample	MNL	[0.161, 0.335]	0	[1.148, 1.319]	[1,228, 2,749]
County nests; full sample	Nested	[0.115, 0.440]	[0.830, 0.943]	[0.439, 1.189]	[156, 288]
County nests; 2015–2016	MNL	[0.127, 0.483]	0	[0.706, 1.016]	[1,570, 2,029]
County nests; 2015–2016	Nested	[0.117, 0.422]	[0.763, 0.941]	[0.384, 1.551]	[244, 810]
Population in demand ; full sample	MNL	[0.150, 0.337]	0	[1.128, 1.333]	[2,591, 3,342]
Population in demand ; full sample	Nested	[0.120, 0.447]	[0.446, 0.986]	[0.517, 1.366]	[259, 1,490]
Population in demand ; 2015–2016	MNL	[0.160, 0.482]	0	[0.761, 1.178]	[2,679, 4,435]
Population in demand ; 2015–2016	Nested	[0.120, 0.402]	[0.435, 0.999]	[0.453, 1.790]	[649, 2,533]
City nests; full sample	Nested	[0.127, 0.387]	[0.212, 0.421]	[0.774, 1.279]	[950, 1,136]
City nests; 2015–2016	Nested	[0.114, 0.779]	[0.247, 0.536]	[0.536, 0.924]	[934, 2,912]

Notes: Each bracket gives the minimum and maximum across the models in the row; brackets are not confidence intervals. MNL fixes $\sigma = 0$.

Table 19: Own-price elasticities in 2016. The store elasticity is the median across legal stores for a price change by that store alone. The legal-sector and illegal-sector elasticities are for a common price change by every seller in the sector, and the all-cannabis elasticity is for a common change in legal and illegal prices.

Specification	Demand model	Store (median)	Legal sector	Illegal sector	All cannabis
County nests; full sample	MNL	[−4.10, −1.97]	[−3.07, −1.48]	[−1.53, −1.04]	[−0.99, −0.67]
County nests; full sample	Nested	[−78.48, −19.51]	[−4.04, −1.06]	[−2.02, −1.51]	[−1.30, −0.98]
County nests; 2015–2016	MNL	[−5.91, −1.55]	[−4.43, −1.16]	[−2.21, −0.58]	[−1.43, −0.37]
County nests; 2015–2016	Nested	[−83.13, −6.09]	[−3.87, −1.08]	[−1.93, −0.56]	[−1.25, −0.36]
Population in demand ; full sample	MNL	[−4.12, −1.83]	[−3.09, −1.38]	[−1.54, −0.70]	[−1.00, −0.45]
Population in demand ; full sample	Nested	[−332.03, −6.25]	[−4.10, −1.10]	[−2.05, −1.34]	[−1.32, −0.86]
Population in demand ; 2015–2016	MNL	[−5.89, −1.95]	[−4.42, −1.47]	[−2.21, −0.96]	[−1.43, −0.62]
Population in demand ; 2015–2016	Nested	[−4,613.36, −3.55]	[−3.68, −1.10]	[−1.84, −0.77]	[−1.19, −0.50]
City nests; full sample	Nested	[−7.16, −1.86]	[−3.55, −1.17]	[−1.77, −0.98]	[−1.15, −0.63]
City nests; 2015–2016	Nested	[−17.53, −2.02]	[−7.15, −1.04]	[−3.57, −0.52]	[−2.31, −0.34]

Notes: Each bracket gives the minimum and maximum across the models in the row; brackets are not confidence intervals. Store elasticities grow without bound as σ approaches one, which explains the largest entries in the population-in-demand rows.

Table 20: Quantity responses to common price increases in 2016, as elasticities. Each panel names the prices that rise; each column names the quantity that responds.

Specification	Demand model	Illegal quantity	Cannabis quantity	Outside quantity
All legal prices rise; illegal prices fixed				
County nests; full sample	MNL	[0.74, 1.10]	[−0.47, −0.32]	[0.70, 1.04]
County nests; full sample	Nested	[1.08, 1.44]	[−0.61, −0.46]	[1.02, 1.36]
County nests; 2015–2016	MNL	[0.41, 1.58]	[−0.67, −0.18]	[0.39, 1.50]
County nests; 2015–2016	Nested	[0.40, 1.38]	[−0.59, −0.17]	[0.38, 1.31]
Population in demand ; full sample	MNL	[0.50, 1.10]	[−0.47, −0.21]	[0.48, 1.04]

Continued on next page

Table 20 (continued)

Population in demand ; full sample	Nested	[0.96, 1.46]	[−0.62, −0.41]	[0.91, 1.39]
Population in demand ; 2015–2016	MNL	[0.69, 1.58]	[−0.67, −0.29]	[0.65, 1.49]
Population in demand ; 2015–2016	Nested	[0.55, 1.32]	[−0.56, −0.24]	[0.52, 1.24]
City nests; full sample	Nested	[0.70, 1.27]	[−0.54, −0.30]	[0.66, 1.20]
City nests; 2015–2016	Nested	[0.37, 2.55]	[−1.09, −0.16]	[0.35, 2.41]

Illegal prices rise; legal prices fixed

Specification	Demand model	Legal quantity	Cannabis quantity	Outside quantity
County nests; full sample	MNL	[0.78, 1.16]	[−0.52, −0.35]	[0.79, 1.16]
County nests; full sample	Nested	[1.14, 1.52]	[−0.69, −0.52]	[1.14, 1.53]
County nests; 2015–2016	MNL	[0.44, 1.67]	[−0.76, −0.20]	[0.44, 1.68]
County nests; 2015–2016	Nested	[0.43, 1.46]	[−0.66, −0.19]	[0.43, 1.46]
Population in demand ; full sample	MNL	[0.53, 1.17]	[−0.53, −0.24]	[0.53, 1.17]
Population in demand ; full sample	Nested	[1.01, 1.55]	[−0.70, −0.46]	[1.01, 1.55]
Population in demand ; 2015–2016	MNL	[0.73, 1.67]	[−0.75, −0.33]	[0.73, 1.67]
Population in demand ; 2015–2016	Nested	[0.59, 1.39]	[−0.63, −0.26]	[0.59, 1.39]
City nests; full sample	Nested	[0.74, 1.34]	[−0.61, −0.33]	[0.74, 1.34]
City nests; 2015–2016	Nested	[0.39, 2.70]	[−1.22, −0.18]	[0.39, 2.70]

All legal and illegal prices rise together

Specification	Demand model	Legal quantity	Illegal quantity	Outside quantity
County nests; full sample	MNL	[−1.91, −1.30]	[−0.44, −0.30]	[1.49, 2.20]
County nests; full sample	Nested	[−2.52, −1.89]	[−0.57, −0.43]	[2.17, 2.89]
County nests; 2015–2016	MNL	[−2.76, −0.72]	[−0.63, −0.17]	[0.83, 3.17]
County nests; 2015–2016	Nested	[−2.41, −0.70]	[−0.55, −0.16]	[0.81, 2.77]
Population in demand ; full sample	MNL	[−1.93, −0.88]	[−0.44, −0.20]	[1.01, 2.21]
Population in demand ; full sample	Nested	[−2.56, −1.67]	[−0.58, −0.38]	[1.92, 2.94]

Continued on next page

Table 20 (continued)

Population in demand ; 2015–2016	MNL	[−2.75, −1.20]	[−0.63, −0.27]	[1.38, 3.17]
Population in demand ; 2015–2016	Nested	[−2.29, −0.97]	[−0.52, −0.22]	[1.11, 2.64]
City nests; full sample	Nested	[−2.21, −1.22]	[−0.50, −0.28]	[1.40, 2.54]
City nests; 2015–2016	Nested	[−4.45, −0.65]	[−1.02, −0.15]	[0.75, 5.12]

Notes: Each bracket gives the minimum and maximum across the models in the row; brackets are not confidence intervals. Other prices and the number of illegal sellers stay fixed. Cannabis means legal plus illegal consumption. Outside means no cannabis purchase.