A Study of Tax Invariance

Benjamin Hansen[*], University of Oregon, NBER, IZA
Kendall Houghton[*], U.S. Census Bureau
Keaton Miller[*], University of Oregon
Caroline Weber[*], University of Kentucky

January 2022

Abstract

We provide an empirical test of tax invariance (TIV). When a 25 percent tax remitted by cannabis manufacturers was eliminated in Washington state and the retail excise tax was simultaneously increased from 25 to 37 percent—a shift intended to be revenue-neutral—TIV did not hold. Manufacturers kept two-thirds of their tax savings instead of passing all their savings through to retail firms as predicted by TIV. Retailers passed one-third of their tax increase on to consumers instead of keeping prices constant or even declining as predicted by TIV.

JEL Codes: H20, H30, H70.

Keywords: Tax Invariance, Natural Experiment, Excise Taxes, Cannabis, Tax Incidence.

[*]Hansen: bchansen@uoregon.edu; Houghton: kendall.a.houghton@census.gov; Miller: keatonm@uoregon.edu; Weber: caroline.weber@uky.edu. The authors would like to thank David Agrawal, Nathan Anderson, Youssef Benzarti, Scott Cunningham, David Evans, Naomi Feldman, Michael Grossman, Bill Hoyt, Donald Kenkel, Michael Kuhn, Nathan Seegert, Juan Carlos Suárez Serrato, Joel Slemrod, Dave Wildasin, and anonymous referees for helpful comments. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Houghton completed substantive work on this paper while at the University of Oregon, before employment at the U.S. Census Bureau. Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. We appreciate comments and feedback from participants at seminars at Case Western, Columbia, Cornell, Norwegian School of Economics, International Online Public Finance Seminar, Portland State, Rutgers, University of Kentucky, and conference participants at ASSA, IHEA, NTA, WEAI, and the IIOC meetings, as well as industry participants and Cannabis Science and Policy Summit attendees. Many thanks to David Shi for excellent research assistance. This paper previously circulated as "Getting into the Weeds of Tax Invariance." Some of the results in this paper previously circulated as part of "The Taxation of Recreational Marijuana: Evidence from Washington State" and some of our thanks are for comments provided on that work.

1 Introduction

Tax invariance (TIV) is a theoretical result that who remits taxes does not influence incidence. It allows policymakers to focus on administrative and evasion costs without considering welfare effects of alternative tax collection strategies. While empirical work suggests that TIV fails under specific circumstances—when evasion opportunities vary along the supply chain (Slemrod, 2008; Kopczuk et al., 2016; Brockmeyer and Hernandez, 2016), when prices are rigid (Muysken et al., 1999; Saez et al., 2012; Lehmann et al., 2013), or if tax salience differs between consumers and firms (Chetty et al., 2009; Finkelstein, 2009)—it is unclear whether TIV does not hold in general, or just that it cannot be applied in particular settings.

We examine a test of TIV that occurred in the cannabis market in Washington state. The specific circumstances under which TIV has been documented to fail do not apply here. The frequently-audited comprehensive regulatory reporting system makes tax evasion difficult. Prices change often, implying rigidities are unlikely. Tax salience is likely high for manufacturers and retailers. Regulatory requirements ensure that owners are well-capitalized. Posted retail prices include all taxes, ensuring consumer salience. Finally, tax leakage and competition are not relevant as the market is closed: each product purchased in Washington was produced in Washington (and vice versa).[1]

Prior to July 1, 2015, a 25% gross receipts tax applied to each transfer in the market. Producers remitted the tax when they sold to manufacturers, manufacturers were taxed on sales to retailers, and retailers were taxed on sales to consumers. After July 1, the retail tax was 37% and all other taxes were eliminated. This change was unexpected by market

[1]Neighboring states did not have legal cannabis markets during our study period.

participants – the reform was passed on June 27 and signed by the Governor on June 30 (La Corte, 2015).

We measure the effects of this reform empirically using an interrupted time series regression in first differences; that is, we ask how prices change in the week of the reform relative to weeks surrounding the reform. Identification rests on the assumption that, after controlling for product characteristics, prices would not have changed in the week of the reform (relative to a baseline trend) if the reform had not occurred. We conduct event studies and placebo permutation tests which provide no evidence to reject this assumption.[2]

We compare the results of this exercise to model-derived and model-free benchmarks. Our setting features market power (Mace et al., 2020; Hollenbeck and Uetake, 2021) and percentage taxes. In an appendix, we follow the modelling approach of Weyl and Fabinger (2013) and show that for a wide range of competitive conducts (including common models of imperfect competition), manufacturers respond to the elimination of their tax by passing through their entire savings; thus, manufacturers' prices should decrease 28.7% from pre-reform levels. We derive an alternative benchmark by assuming monotonicity of cost pass-through i.e. by asking how much pass-through from manufacturers would be required to maintain constant tax-inclusive retail prices. Since the reform slightly decreased the total tax burden, manufacturers should decrease their prices at least 17.7% in order to leave retailers' per-gram profits and consumer-facing tax-inclusive prices constant. We find that manufacturers reduce their prices by only 7.2% and therefore reject the null hypothesis of TIV based on either benchmark at the 0.1% level.

[2]We employ this approach rather than a difference-in-differences design as the only potential comparison state is Colorado, which had a significantly different regulatory and industry structure—the assumption that prices in the two states co-move in the study period is likely much stronger than the assumptions we impose.

We then examine retail behavior. Our model framework predicts that retailers should either leave tax-inclusive prices constant or decrease them (depending on the competitive environment). Instead, we find tax-inclusive retail prices *increased* by an average of 2.5%. Retailers pass-through one-third of the tax increase. Another roughly one-third is borne by manufacturers, leaving retailers with about one-third of the increase. We find evidence that retailers maintained constant tax-exclusive markups, consistent with our model's pricing rule conditional on manufacturers' behavior.

In summary, our results are inconsistent with TIV: a reform that should not have changed the welfare of manufacturers, retailers, and consumers instead increased the profits of manufacturers at the expense of retailers and consumers. We conclude by discussing potential mechanisms for this result, implications for policymakers, and future research.

2 Background

We analyze the adult-use cannabis market in Washington state, which opened in July 2014. We have written elsewhere about the history of this market (Hansen et al., 2020; Miller and Seo, 2021)—here we focus on features of the market and the reform underpinning our analysis.

The market consists of three types of firms: cultivators, who grow cannabis plants, manufacturers, who transform plant material into marijuana products, and retailers, who sell products they obtain from manufacturers to consumers. Potential entrants have to pass background checks and undergo a lengthy regulatory process requiring substantial capital investment before entry. Cultivators face capacity constraints—the largest firms may cultivate 30,000 sq. ft. of plant canopy and may not merge to increase capacity. While retailers must be financially independent from other firms, a cultivator and a manufacturer may vertically integrate. When the reform was implemented, approximately 94% (by weight) of usable marijuana—dried and cured cannabis flowers—was produced through a vertically-integrated process (Hansen et al., 2020b). Thus, we focus our analyses on two types of firms, "manufacturers" and "retailers".[3]

The market is closed: all cannabis sold by retailers is grown in the state, and every ounce grown legally within the state is sold at a Washington retailer. These rules are enforced through the state's "seed-to-sale" traceability system, which tracks each plant from cultivation through processing and retail. The collected data can be used to check for tax evasion: retailers cannot sell cannabis without manufacturing records, which forces manufacturers to report accurately.[4] Reporting is enforced through frequent audits—firms typically face one or more visits per year—backed by significant penalties.[5]

Washington's initial tax regime consisted of a 25% tax collected at every transfer of cannabis. Vertically-integrated manufacturers owed no tax on intra-firm transfers. The reform we analyze eliminated the 25% excise taxes within the supply chain and increased the retail excise tax rate to 37%. The excise tax at retail applied to the sales-tax-inclusive price pre-reform and the sales-tax-exclusive price post-reform. Accounting for changes to the

[3]State law calls cultivators "producers" and manufacturers "processors"—we choose nomenclature to represent functional equivalents in other markets.

[4]Retailers can under-report sales, but such behavior is detectable by comparing retail sales and wholesale purchases. Our estimates are unaffected by dropping the few retailers that under-report substantially.

[5]We observe audit violations and our estimates remain approximately the same if we drop firms with under-reporting violations. Moreover, our results are not consistent with asymmetric tax evasion. Manufacturers would pass along less tax savings if retailers are more effective tax evaders, but in that case, retailers would still not pass along any of the tax to consumers.

base and rate of the retail tax, the reform changed the average retail tax rate by 6.93%.[6] This change was designed to be revenue neutral under the assumption that tax-*exclusive* prices remained constant (whereas TIV predicts constant tax-*inclusive* prices). We account for both the change in the rate and the base of the retail excise tax in our analyses. We provide calculations of revenue pre- and post-reform in Section 3. Other regulations concerning cannabis production, distribution, and sales were unaffected.

Our identification assumes that the policy change was unanticipated. The bill originated and was passed in the Washington House during the 2015 Regular Session, but stalled in the Senate. The bill was reintroduced in the First Special Session, but again stalled. Finally, on June 27 during the Second Special Session, the bill passed both chambers. The Governor signed it on June 30 and the law went into effect the next day. Contemporaneous reporting portrayed the industry as unprepared. According to one retail manager, "[we] have a few hours to change an entire market's pricing structure. It is an exceptionally short window for such a tremendous change" (La Corte, 2015).

3 Data and Methods

Our data consist of records from the "traceability" system maintained by the Washington State Liquor and Cannabis Board (WSLCB). We obtain data on all plants, products, and sales. We restrict our analysis to "usable marijuana" products—74.5% of the total transactions observed in our data. Within this category, products are differentiated by "strain" (analogous to fruit cultivars), potency, and form-factor (loose or pre-rolled). These charac-

[6] The average sales tax rate during this period was 8.9%, thus $\log\left(\frac{1.25(1+.089)}{1.37+.089}\right) = -0.0693$

teristics are captured by our fixed effects.

We aggregate retail sales by inventory-lot-week, where an inventory-lot is a batch of identical product. We exclude firms with less than two months of pre- and post-reform data. The reform also changed technical reporting rules which affect the price data. We clean the data to reflect the prices faced by consumers using an algorithm based on retail rounding behavior verified by spot checks of historical menus.[7] See Appendix C for details.

Table 1 reports summary statistics for retail lots for the six weeks pre-reform (the basis for our analyses in Section 4). The average tax-inclusive retail price was $13.03 per gram and the tax-exclusive price was $9.57 per gram. Retailer tax-exclusive prices are more than double manufacturer tax-inclusive prices.

The average market share of retailers in the 10-mile radius around their location was 31%, suggesting that there is substantial retail market power (Mace et al., 2020; Hollenbeck and Uetake, 2021). The manufacturer market is effectively state-wide; the average market share is 1.4% and none exceeds 7%. However, manufacturers may exert market power on individual retailers. We construct a retail-level manufacturer concentration index by sorting each retailer's suppliers by the weight of inventory sold and counting the number of manufacturers that comprise at least 75% of sales. On average, seven manufacturers supply 75% of a retailer's inventory.

[7]Cannabis retailers set tax-inclusive prices that are round numbers (e.g. $8 or $10.25). While this represents a potential friction, the effective minimum price change is smaller than the effects we estimate.

3.1 Theoretic Benchmarks

As TIV is not axiomatic – it is derived from models of supply and demand behavior – it is necessary to establish benchmarks against which to compare our estimates of price changes pre- and post-reform. We establish two such benchmarks. The first is inspired by the observation that the supply chain for cannabis is imperfectly competitive, with substantial market power both at the retail and at the manufacturer level. In Appendix A, we write a model of the supply chain in the style of Spengler (1950) and Weyl and Fabinger (2013) that nests popular models of imperfect competition including homogeneous products oligopoly, differentiated Nash-in-prices, and monopolistic competition. We show that manufacturers should pass-through tax changes to retailers regardless of the competitive environment. In this case, our framework predicts that manufacturers should decrease their tax-inclusive prices by $\log(1 - 0.25) = -28.7\%$. If we estimate a different price response to the reform, we can reject the (very general) assumptions of the framework.

Despite the broad flexibility of the Weyl and Fabinger (2013) structure, it may be possible to construct models which a) rationalize other price responses and b) feature a TIV result. We therefore construct an alternative benchmark by asking what we would learn about responses to this reform from models that satisfy weak monotonicity in cost pass-through and provide a TIV result. If TIV were to hold and a revenue-neutral reform occurred, tax-inclusive retail prices would remain constant, as would post-tax retail and manufacturers' profits. Washington's reform decreased the total tax burden per gram, which implies that firms should respond by *at least* holding tax-inclusive retail prices and post-tax profits constant (or lowering retail prices), as not doing so would imply non-monotonicity in cost pass-through. For example, a model of Nash bargaining between manufacturers and retailers may violate the assumptions of Appendix A yet satisfy TIV. If costs decrease through a tax rate reduction, surplus increases and bargaining participants should not be made worse off i.e. costs are passed-through weakly monotonically.

Given pre-reform prices, to maintain a constant tax-inclusive retail price and constant per-gram retail profits (and therefore to satisfy TIV), manufacturers would have to decrease their prices by an average of 64 cents, or 17.7%[8] This alternative benchmark therefore represents the most conservative notion of behavior that could be compliant with TIV across models that feature monotonicity in cost pass-through, a weaker assumption than that imposed by the framework of Appendix Section A.[9]

Under a revenue-neutral reform, TIV predicts that retailers would reduce their tax-*exclusive* prices by 6.93% (the change in the retail tax rate) and maintain constant tax-*inclusive* prices. Under weak substitution and imperfect competition, we predict tax-inclusive prices will decline. As we calculate the reform is slightly revenue-decreasing, our framework suggests retailers should reduce tax-inclusive prices further.

To summarize, if we estimate a decrease in average manufacturer tax-inclusive prices of less than 28.7%, we reject our framework and therefore the TIV results of Weyl and Fabinger (2013). If we estimate a decrease in average manufacturer tax-inclusive prices of less than 17.7%, we reject TIV under any model with monotone cost pass-through. Moreover, if we estimate any *increase* in retailer tax-inclusive prices, we reject both our framework and the

[8] 13.03/(1.37+0.089)-13.03/(1.25*(1+0.089)) = 64 cents.

[9] Models of sequential entry and capacity setting can feature non-monotone policy functions driven by entry foreclosure incentives (see e.g. Cabral, 2017). In our context, entry was restricted by the government and lotteries were used to allocate licenses (Thomas, 2018). Differences in entry timing were likely driven by local regulatory processes and it is therefore unlikely that entry foreclosure incentives were first-order.

more general monotonicity assumption. Since the pre- and post-reform tax rates are constant across firms, these benchmarks can be compared to estimated average treatment effects.

3.2 Empirical Specification

We model responses to the tax reform as an interrupted time-series in first differences:

$$\Delta log(p_{it}) = \alpha_0 + \alpha_1 \Delta TaxReform_t + \alpha_2 FE_i + u_{it}, \tag{1}$$

where i is the unit of observation as described for the manufacturer and retail analyses below, and t indicates the week. p is the wholesale or retail price per gram, $TaxReform$ is an indicator variable that is one after July 1, 2015 and zero before, and FE are fixed effects. α_1 is the parameter of interest.[10] Because all firms are treated and generally included in our analysis, this parameter is an ATT. Our analysis window spans six weeks before and after the reform.[11] We two-way cluster standard errors on manufacturer and retail location (Cameron et al., 2011).[12] Our identifying assumption is that within a given product, there are no shocks in the week of the reform that would have a significant and systematic impact on prices besides the direct effect of the tax reform. Given the short interval between observations (i.e a week, not a year), this assumption is plausible.

Our identifying assumption is much more likely to hold in our setting than the assumption made in a classic difference-in-differences design – that, for a lengthy period (e.g. a year), two states would experience the same systematic price shocks in the cannabis market. In general,

[10] Without fixed effects, this regression is equivalent to an interrupted time series regression in levels with fixed effects at the level of our first differences and a control for time to the reform.

[11] We demonstrate our estimates are robust to this choice in Figure B.5.

[12] Firm clusters or two-way clusters on firm and week yield similar standard errors.

it's not clear which of these two would be more likely to hold *a priori*, but in this setting, it is very unlikely that the latter would hold because each state is a closed market and there are substantial differences in market regulatory structures, processes and outcomes.Moreover, we will provide evidence through placebo analyses and event studies that we have no reason to reject the null hypothesis that our identifying assumption is valid. When we use the same data to examine another natural experiment that allows for the addition of a comparison group to our RDIT design[13] (i.e. a difference-in-discontinuities regression) (Hansen et al., 2020a), we found the estimates remained quantitatively and qualitatively similar.

Our implementation of interrupted time series, also known as regression discontinuity in time (RDIT), addresses critiques previously raised against this method (Hausman and Rapson, 2018). We select a narrow bandwidth (measured in weeks, not years) and we estimate the regression in first-differences rather than levels; this, along with our event study figures, allows us to precisely pin down the response in the week of the reform, rather than allowing the regression to obtain part of its identification from shifts that may happen many weeks away from the reform. In addition, we address autocorrelation over time with firm-level standard error clustering, we aggregate at the weekly level to avoid challenges associated with estimating day-of-week fixed effects, and we include fine-grained fixed effects to address any compositional shifts over time.

For the manufacturer analysis, we aggregate to the manufacturer-retailer-strain-week level, so that i is a manufacturer-retailer-strain tuple, and then take first differences.[14] Each manufacturer-retailer-strain tuple does not sell every week. We thus calculate the minimal-

[13] Washington firms near the Oregon border were "treated" while those away from the border were not.

[14] Aggregation beyond the inventory lot is required because each lot is sold only once. Other possible aggregations produce similar estimates with lower power (though statistical significance remains).

length difference and include difference-length fixed effects.[15] The maximum difference-length allowed is 4 weeks.We are thus estimating the magnitude of price changes in response to the reform within a specific firm-product pairing. Retailer-manufacturer-strain fixed effects allow each retailer-manufacturer-strain to have a separate time trend.

For our main retail analysis, we aggregate to the inventory-lot-week level so that i is a retail inventory lot. Retail sales from a given inventory lot are frequent, so we construct one-week differences. We are thus estimating the change in the retail price of an inventory lot in response to the tax reform holding all possible product and firm variation constant. Sales of retail inventory lots typically last multiple weeks, so we include fixed effects for weeks from the first sale out of that inventory lot. Inventory-lot fixed effects allow prices in each inventory-lot to have a separate time trend.

We separately examine the first week of retail sales for each inventory-lot and include only those that were purchased from manufacturers in the same week. Similar to our manufacturer analysis, we aggregate by retailer-manufacturer-strain and take varied length differences. We include difference-length fixed effects. In these regressions, we ask how prices for *new inventory-lots purchased post-reform* change relative to *pre-reform lots of the same strain from the same manufacturer*. This allows us to examine whether prices change more or less if the inventory was purchased post-reform relative to inventory that had already been purchased and was selling pre-reform.

[15]These fixed effects are not significant. Our estimates are similar when restricted to one-week differences, but with less power.

4 Results

4.1 Manufacturer Price Response

Table 2 reports estimates of Equation (1) for manufacturers. Column (1) estimates a 6.5% price decline in response to the tax reform. When we add manufacturer-retailer-strain fixed effects in Column (2) – our baseline specification – the point estimate becomes -7.2% (p-value: 0.000). This is roughly one-third of the 17.7% price decrease needed to preserve retailer per-gram profits (and therefore to minimally satisfy TIV), and one-quarter of the 28.7% decrease predicted by our framework. We can reject the null hypothesis that our estimate is consistent with TIV (p-value: 0.000). Column (3) repeats Column (2) for the price in levels instead of logs – the reform decreased manufacturer prices by 23 cents, about one-third of the 64 cent benchmark. Table 2 Column (4) aggregates the data by months instead of weeks and we find the estimates are very similar with smaller standard errors. This aggregation is an alternative way to address manufacturers not selling every strain they produce to every retailer every week, and also allows us to examine a longer-term response.

The bottom panel of Table 2 repeats the specification of each column for a placebo reform dated one year later. The estimates are near zero across all four columns, providing support that our regression specifications are valid.[16] One could subtract the placebo estimates from the main estimates to create a differences-in-RD design; the estimates would be very similar.

It is possible that our rejection of TIV is driven by price stickiness or that the dampened response we observe (relative to the TIV benchmark) is driven by the sheer magnitude of

[16] In addition, we estimate many placebo regressions and construct a placebo permutation test for manufacturers in the top panel of Figure B.4. There are no placebo estimates as extreme as our estimate and the implied p-value is 0.012.

the TIV-consistent response relative to typical price shifts in this market. Figure 1 provides evidence of price mobility by plotting the entire distribution of weekly price changes for each retail-manufacturer-strain pair in the baseline estimate as a histogram in the top panel (and monthly price changes in the bottom panel). The period of the tax reform is marked by the hollow green histogram and surrounding periods are marked by the gray histogram. The width of each bin is 0.04, so all price changes within 2% of zero are included in the bin centered around zero. The dashed lines from right to left indicate: full pass-through predicted by models of firm behavior (-28.7%), our alternative benchmark of pass-through (-17.7%), and zero (i.e no price change).

These histograms illustrate several things. First, we see the shift in the distribution in the period of the reform. Note that there is a substantial increase in pass-through all the way down to the full pass-through benchmark. While TIV fails in aggregate, there are retail-manufacturer-strain tuples for which TIV appears to hold in isolation. Second, the histograms highlight that large price shifts do occur with reasonable frequency in the absence of the reform – about 16 percent of retail-manufacturer-strain tuple weekly price changes were at least as large as our alternative benchmark. This statistic rises to almost 20 percent at the monthly level. Moreover, less than 10 percent of retail-manufacturer-strain tuples do not adjust their prices in the period of the reform. Even if we rescaled our estimate assuming that any observation with minimal adjustment in this period was caused by rigidities or lack of awareness, the data would still reject the null hypothesis of TIV.[17]

Figure 1 is consistent with our identification assumptions and therefore provides addi-

[17]However, this does not rule out additional adjustments by many firms over a longer horizon. To fully assess whether there is evidence of a longer-run adjustment towards TIV, we consider event studies in Section 4.3.

tional evidence of the validity of our regression estimates – the price shifts are concentrated in the region one would expect. That is, there are substantial increases in the distribution of price decreases in the period of the reform in each of the bins up to our full pass-through benchmark, and much less beyond that benchmark. As expected, this can be seen most cleanly in the top panel for weekly price changes. If the response we estimate were partially attributable to some more generic market shift in that same week, there is no reason to think that the price shifts would have this particular pattern.

4.2 Retail Price Response

Table 3 reports estimates of Equation (1) for retailers. The coefficients are very similar in Columns (1) and (2). The coefficient in Column (2) – our baseline specification – implies that the reform reduced tax-exclusive retail prices by 4.4% (p-value: 0.000). Combined with the rate change, this implies that tax-*inclusive* prices increased by 2.5%; retailers passed through roughly one-third of the tax to consumers. We can reject the null hypothesis of TIV-consistent pricing behavior (p-value: 0.000).

As firms might have taken time to adjust (and the Independence Day holiday may have generated temporary price adjustments), Column (3) repeats Column (2) for two week differences and drops the first week after the reform, so that the effect of the reform is identified from the difference between the week before and the week after the reform. The estimates are approximately the same, indicating that neither of these concerns play a large role. We will return to a broader discussion of timing in Section 4.3.

Table 3 Column (4) repeats Column (2) with the dependent variable in levels—we estimate that average retail tax-exclusive prices fell by 41 cents per gram. Thus, retailers are an average of 41 cents per gram worse off on existing inventory as a result of the reform. On fresh inventory, retailers were roughly 18 cents per gram worse off (41 less the estimated 23 cent decrease in manufacturer prices estimated in Table 2). In other words, under TIV this reform should have caused manufacturer and retail tax-exclusive prices to fall by 64 cents leaving profit and consumers unaffected. Instead, it caused smaller manufacturer price cuts leaving both retailers and consumers worse off.

Table 3 Columns (5) and (6) take an alternative approach to identification examining inventory lots only in their first week and only if retailers purchased the inventory lot from the manufacturer in that week. For this, we create a panel of retail-processor-strain-weight group-weeks. The estimates are quite similar suggesting retailers' price responses are largely unaffected by whether they are selling inventory lots purchased pre-reform or selling new inventory lots purchased post-reform. Column (6) adds the first-differenced log manufacturer price; the coefficient on the wholesale price is not statistically different from one and the coefficient on $\Delta TaxReform$ is now approximately zero. This suggests retailers largely maintained a constant tax-exclusive markup.[18] This is consistent with the adjusted Lerner pricing rule of Equation (2). Hence, while retail behavior *as a whole* is inconsistent with TIV, after conditioning on the pass-through from manufacturers, retailer behavioris consistent with marginal-cost pricing (and therefore potentially consistent with TIV).

In the bottom panel of Table 3 and in the bottom panel of Figure B.4, we repeat the placebo analyses we perform for our manufacturer sample for the retailer sample. These

[18]We could also estimate the response of retail margins to the tax reform directly on this sample to reach the same conclusion.

analyses provide no evidence of bias in our estimates.

4.3 Event Studies

The analyses above indicate that prices changed at the time of the reform—yet it is possible that these changes were part of the ongoing evolution of the market. Moreover, the estimates above do not indicate whether there is additional adjustment towards TIV beyond the first week. To address these issues, we conduct event studies for both the manufacturer and retailer responses using our baseline specifications from Tables 2 and 3. For manufacturers, we do not drop the $t-1$ tax reform coefficient due to our varied difference lengths.[19] Figure 2 plots the relevant coefficients and confidence intervals.[20]

Figure 2 demonstrates no discernible trend in prices pre-reform. This implies that once we control for the compositional shifts in Figure B.1 with appropriate fixed effects, we no longer observe any significant trends in prices prior to the reform. The entire response happens in period t, the reform week. Given the varied difference lengths for manufacturers, this implies that manufacturers adjust their prices the first time they sell a particular retail-strain pair post-reform. This is compelling evidence that our estimates are unlikely to be driven by any ongoing market evolution and are instead a true response to the reform. The immediate nature of the response suggests that prices in this market follow a unit root process, further supporting our first-difference specification. Moreover, this suggests that our results are not driven by learning or other short-run adjustments. Figure B.3 provides longer-run evidence by repeating the manufacturer event study aggregated at the monthly level. We find no

[19]E.g., for a two week difference that spans t-1 to t+1, both the t and t+1 coefficients are relevant.

[20]Figure B.2 replicates the event study plots one year later, further emphasizing the placebo findings in previous sections—our identification strategy is effective in this setting when tested in other periods with similar cyclicality and holiday patterns.

evidence of long-run learning either (Doraszelski et al., 2018; Huang et al., 2018).[21]

5 Discussion and Conclusion

TIV is a key component of tax policy design and analysis—it states that taxes may be collected at any point in the supply chain without concern as to the ultimate incidence. While the literature has documented cases in which TIV failed to hold, these results have come with caveats driven by specific frictions or asymmetries present in the markets studied. We study a reform in a market with none of the these issues and show that TIV did not hold: a reform intended to be welfare-neutral or even welfare-enhancing had negative consequences for both retailers and consumers. This result is driven by manufacturers, who on average reduced prices significantly less than broad classes of models consistent with TIV would predict. Conditional on manufacturer prices, we find evidence that retailers applied a constant markup over marginal costs, consistent with standard models of firm behavior.

Our empirical results allow us not only to reject the benchmark of our framework, which nests standard market power explanations, but also a more conservative model-free benchmark derived from monotonicity of cost pass-through, and thus these results are likely robust to a great deal of model uncertainty. In particular, if manufacturers employed average-cost pricing mechanisms (Hall and Hitch, 1939; Altomonte et al., 2015), we would expect the reform to cause similar or larger price drops than under marginal-cost pricing. While the reform eliminated incentives for inefficient vertical integration and, in the long run, production

[21] Additionally, Figure B.1 plots matched retail and manufacturer prices in their first week of retail sale pre- and post-reform. If there were a slow-moving adjustment towards TIV, the gap between matched wholesale and retail prices (which will both experience the same market level shocks) would slowly diverge as wholesale prices continued to adjust downwards relative to retail prices towards the TIV equilibrium. We see no evidence of this behavior.

increased (Hansen et al., 2020b); increased production efficiency should similarly drive down prices. The frequency of price changes—and the prevalence of at least some drop in manufacturer prices in response to the reform—suggest that managerial inattention is not relevant (Gabaix, 2019). Our event studies suggest the response is immediate, which decreases the likelihood that learning can explain our findings.

We thus turn to the literature establishing asymmetric firm responses to changes in market conditions for possible explanations, as others have found asymmetric firm behaviors in related settings. The work of Benzarti et al. (2020) is particularly relevant—they find increases in value-added taxes are passed-through to consumers at twice the rate of decreases. In our setting, retailers, which experienced a tax increase, passed-through taxes in a way that is consistent with standard models of profit maximization, while manufacturers, which experienced a tax decrease, failed to pass-through their savings as predicted. Unlike the VAT context, however, our setting features a simultaneous change and a marketplace where firms and consumers are highly aware of relevant prices; furthermore both retailers and manufacturers engage in repeated transactions with each other over a long period of time. More broadly, the industrial organization literature has identified potential asymmetries in firm responses to changes in demand and costs (Butters et al., 2019; DellaVigna and Gentzkow, 2019).

We view our results as consistent with models that generate asymmetric responses to changes in market conditions due to behavioral phenomena, as opposed to information, transaction, or competitive frictions. In particular, anchoring and loss aversion may explain the outcomes we observe (Kahneman et al., 1982, 1991; Bernheim and Rangel, 2009). While the modal response by manufacturers in the week of the reform was to adjust their prices, the default option of "doing nothing" by maintaining constant tax-inclusive manufacturer prices (and thus realizing a significant increase in variable per-unit profits) may have anchored their negotiations with retail firms. The relatively common and small changes in manufacturer prices we do observe may be a result of competition—manufacturers may "do something" if they incorporate quantity or reputation effects into their analysis of post-tax outcomes (Rotemberg, 2011) and competitors may be compelled to act as a consequence. In contrast, in aggregate, retailers may have overcame their default "do nothing" option (constant tax-inclusive prices) because this option represented a loss in variable per-unit profit. Once the default was overcome, they made decisions consistent with standard models.

Our findings have wide-ranging implications for tax policy. First, designers of new taxes may face welfare tradeoffs when choosing where in a supply chain to locate a tax. Both efficiency and equity considerations arise. When considering efficiency, variation in elasticities or competitive structures across the market may affect optimal tax placement. In terms of equity, if a policy goal is to ensure all market participants bear portions of the tax, it may be necessary to impose taxes on these different groups directly.

Second, policymakers considering changes to existing tax policy may face greater consequences for doing so. While it may be possible to implement revenue-neutral reforms, restructuring could create winners and losers. In this case, manufacturers benefited—despite being in an arguably more competitive market—while retailers and consumers were harmed.

Taken together, these concerns point to broader political economy issues surrounding tax policy (Winer and Hettich, 2006; von Hagen, 2008). Political systems may be designed to limit the ability of policymakers to enact tax reforms and thus rational actors may unknowingly design systems which have additional inefficiencies as described here. Indeed, in

Washington state, the legislature may not reform measures passed by ballot initiative for two years after passage. Though local government officials knew from the moment of passage that the gross receipts tax was likely to have negative consequences on the market, their hands were tied. Flexibility in political and policy systems may help avoid these concerns—though at the cost of volatility and asymmetric responses.

Finally, these results demonstrate a need for further experimental and modelling work. Modern models of competition, growth, trade, inflation, and the business cycle generally make assumptions about taxes which are appealing from a tractability standpoint. These assumptions generally imply TIV (e.g. Judd, 2002; Melitz, 2003; Galí, 2015). Our work provides more evidence that policymakers may not be able to rely on TIV because of the ubiquitous nature of default options in reform contexts. In the absence of TIV, it may be necessary to conduct experiments which examine the way in which firms and consumers respond to tax policy and construct models which more accurately capture these responses.

References

Altomonte, C., A. Barattieri, and S. Basu (2015). Average-cost pricing: Some evidence and implications. *European Economic Review 79*, 281–296.

Benzarti, Y., D. Carloni, J. Harju, and T. Kosonen (2020). What goes up may not come down: Asymmetric incidence of value-added taxes. *Journal of Political Economy*, forthcoming.

Bernheim, B. D. and A. Rangel (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics 124*(1), 51–104.

Brockmeyer, A. and M. Hernandez (2016). Taxation, information, and withholding: Evidence from Costa Rica. Working paper.

Butters, A., D. Sacks, and B. Seo (2019). How do national firms respond to local shocks? Evidence from excise taxes. *Kelley School of Business Research Paper*.

Cabral, L. M. (2017). *Introduction to industrial organization.* MIT press.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Chetty, R., A. Looney, and K. Kroft (2009, September). Salience and taxation: Theory and evidence. *American Economic Review 99*(4), 1145–77.

DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in us retail chains. *The Quarterly Journal of Economics 134*(4), 2011–2084.

Doraszelski, U., G. Lewis, and A. Pakes (2018). Just starting out: Learning and equilibrium in a new market. *American Economic Review 108*(3), 565–615.

Finkelstein, A. (2009, 08). E-Z tax: Tax salience and tax rates. *The Quarterly Journal of Economics 124*(3), 969–1010.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework and its applications.* Princeton University Press.

Hall, R. L. and C. J. Hitch (1939). Price theory and business behaviour. *Oxford Economic Papers* (2), 12–45.

Hansen, B., K. Miller, B. Seo, and C. Weber (2020). 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 (2020a). Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. *Journal of Public Economics 187*, 104–159.

Hansen, B., K. Miller, and C. Weber (2020b). Vertical integration and production inefficiency in the presence of a gross receipts tax. Working paper.

Hausman, C. and D. S. Rapson (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics 10*(1), 533–552.

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

Huang, Y., P. B. Ellickson, and M. J. Lovett (2018). Learning to set prices in the Washington state liquor market. *Manuscript. University of Rochester. Simon Business School*.

Judd, K. L. (2002). Capital-income taxation with imperfect competition. *American Economic Review 92*(2), 417–421.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives 5*(1), 193–206.

Kahneman, D., S. P. Slovic, P. Slovic, and A. Tversky (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Keen, M. (1998). The balance between specific and ad valorem taxation. *Fiscal Studies 19*(1), 1–37.

Kopczuk, W., J. Marion, E. Muehlegger, and J. Slemrod (2016, May). Does tax-collection invariance hold? Evasion and the pass-through of state diesel taxes. *American Economic Journal: Economic Policy 8*(2), 251–86.

La Corte, R. (2015, June 30). Gov. Inslee signs recreational marijuana reform bill. *Associated Press*.

Lehmann, E., F. Marical, and L. Rioux (2013). Labor income responds differently to income-tax and payroll-tax reforms. *Journal of Public Economics 99*, 66 – 84.

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

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

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

Muysken, J., T. V. Veen, and E. D. Regt (1999). Does a shift in the tax burden create employment? *Applied Economics 31*(10), 1195–1205.

Rotemberg, J. J. (2011). Fair pricing. *Journal of the European Economic Association 9*(5), 952–981.

Saez, E., M. Matsaganis, and P. Tsakloglou (2012). Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece. *The Quarterly Journal of Economics 127*(1), 493–533.

Skeath, S. E. and G. A. Trandel (1994). A pareto comparison of ad valorem and unit taxes in noncompetitive environments. *Journal of Public Economics 53*(1), 53 – 71.

Slemrod, J. (2008). Does it matter who writes the check to the government? The economics of tax remittance. *National Tax Journal 61*(2), 251–275.

Spengler, J. J. (1950). Vertical integration and antitrust policy. *Journal of Political Economy 58*(4), 347–352.

Suits, D. B. and R. A. Musgrave (1953). Ad valorem and unit taxes compared. *The Quarterly Journal of Economics 67*(4), 598–604.

Thomas, D. (2018). License quotas and the inefficient regulation of sin goods: Evidence from the washington recreational marijuana market. *Available at SSRN 3312960*.

von Hagen, J. (2008). Political economy of fiscal institutions. In D. A. Wittman and B. R. Weingast (Eds.), *The Oxford Handbook of Political Economy.*

Weyl, E. G. and M. Fabinger (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy 121*(3), 528–583.

Winer, S. L. and W. Hettich (2006). Structure and coherence in the political economy of public finance. *The Oxford Handbook of Political Economy 7*, 441.

Tables

Table 1: Pre-Reform Retail Summary Statistics

Variable	**Obs.**	**Mean**	**Std. Dev.**
Prices and Taxes			
Tax-Inclusive Retail Price ($/g)	63,668	13.033	3.798
Tax-Exclusive Retail Price ($/g)	63,668	9.570	2.783
Manufacturer Price ($/g)	63,668	4.103	1.309
Retail State + Local Sales Tax Rate	63,668	1.089	0.006
Tax Revenue Pre-Reform ($/g)	63,668	4.489	1.246
Competition			
Market Share of Retailer in 10 Mile Radius	63,668	0.313	0.282
Market-level Manufacturer Market Share	63,668	0.014	0.016
Retail-Level Manufacturer Concentration Index	63,668	6.997	2.691
Benchmarks Assuming TIV			
Expected Tax Revenue Post-Reform ($/g)	63,668	4.104	1.200
Manufacturer Pass-Through Cents	63,668	−0.640	0.185
Manufacturer Pass-Through Percent Change	63,668	−0.177	0.058

An observation is an inventory-lot-week pre-reform. The data come from our retail analysis set and cover the six weeks prior to the tax reform. Tax revenue is calculated using both excise and state and local sales taxes. The retail-level manufacturer concentration index is calculated as follows: for a given retailer, we sort their suppliers by the weight of inventory sold, and count the number needed to comprise at least 75% of total sales. The "benchmarks assuming TIV" account for changes in the base and rate of the retail excise tax. The "manufacturer pass-through" statistics assume constant tax-inclusive retail prices and indicate the post-reform changes to manufacturer prices that would have left retailer variable-profit-per-gram constant.
†These probabilities are calculated for the subset of retail-processor-strain-weight group-weeks when the inventory lot changes (and thus a new purchase from a manufacturer has occurred).

Table 2: Manufacturer Price Response

	(1) Δlog(Price)	(2) Δlog(Price)	(3) ΔPrice	(4) Δlog(Price)
		Tax Reform		
ΔTax Reform	-0.065^{***}	-0.072^{***}	-0.228^{***}	-0.059^{***}
	(0.015)	(0.018)	(0.068)	(0.014)
Observations	12,087	12,087	12,087	20,902
Manufacturer Firms	199	199	199	210
P-Value for Test of TIV-Predicted Pass-Through	0.000	0.000	0.000	0.000
		Placebo		
ΔPlacebo	0.001	0.000	0.014	-0.002
	(0.012)	(0.014)	(0.040)	(0.010)
Observations	21,288	21,288	21,288	42,354
Manufacturer Firms	180	180	180	208
Bandwidth	6 weeks	6 weeks	6 weeks	6 months
MRS FE?	No	Yes	Yes	Yes
Aggregation	Weekly	Weekly	Weekly	Monthly

This table reports estimates of Equation (1) – other variables in that equation are included, but not reported. An observation is a manufacturer-retailer-strain-week. The outcome is the change in the log of the price per gram charged by the manufacturer to the retailer (except for in column (3) which is the same outcome, but not logged). MRS stands for manufacturer-retailer-strain fixed effects. The estimates are weighted by the total grams sold across the two weeks of the difference. The P-value tests the null hypothesis that the estimated pass-through is equal to that predicted by TIV. For the placebo regressions, we repeat the analysis one year later. These regressions are estimated with `reghdfe` in Stata. Standard errors twoway-clustered by manufacturer and retailer are in parentheses (Cameron et al., 2011). *5% significance level. **1% significance level. ***0.1% significance level.

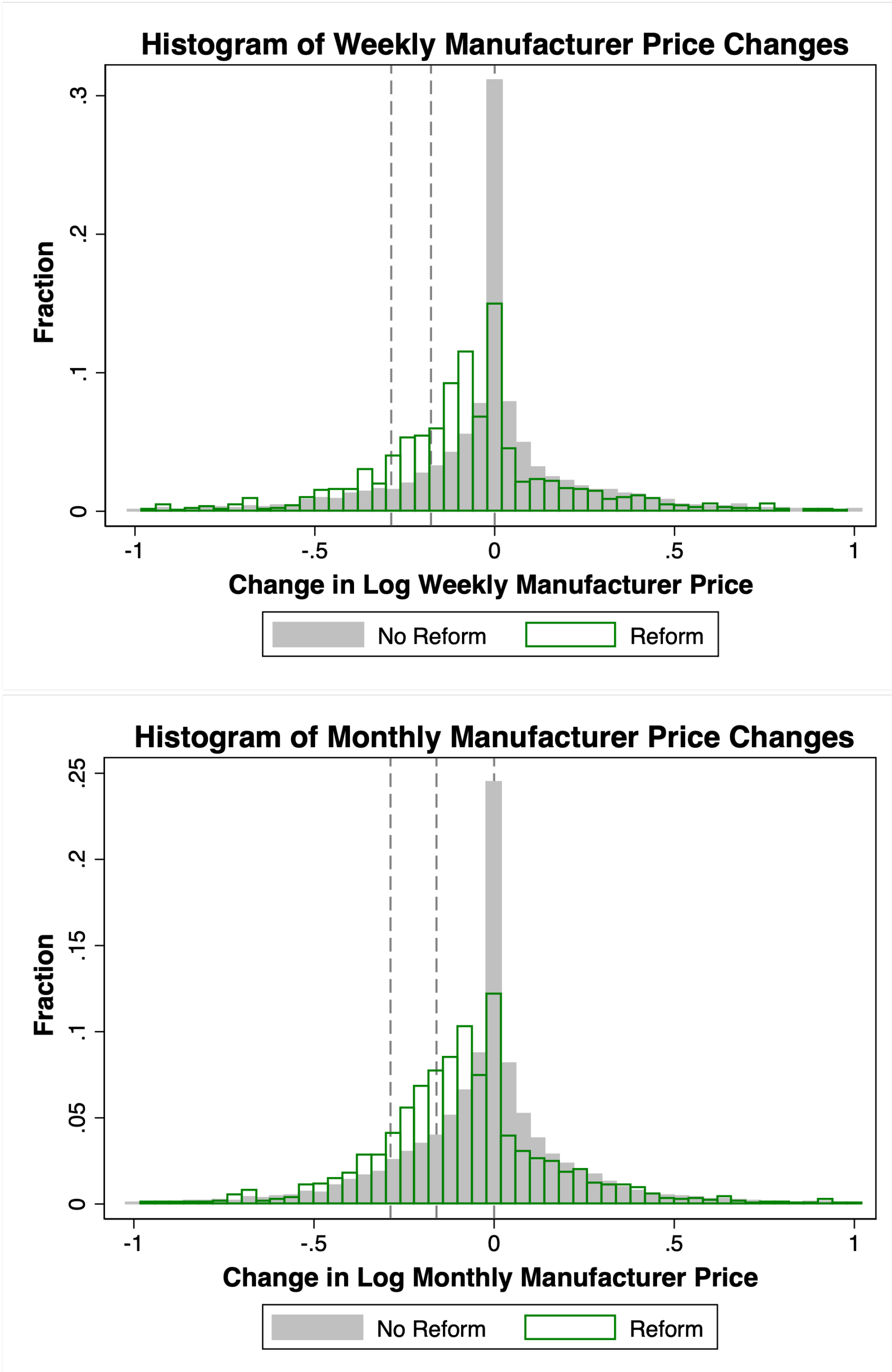
Table 3: Retail Tax-Exclusive Price Response

	(1)	(2)	(3)	(4)	(5)	(6)
	Δlog(Price)	Δlog(Price)	Δlog(Price)	ΔPrice	Δlog(Price)	Δlog(Price)
			Tax Reform			
ΔTax Reform	-0.045^{***}	-0.044^{***}	-0.046^{***}	-0.413^{***}	-0.049^{**}	0.011
	(0.006)	(0.007)	(0.006)	(0.065)	(0.018)	(0.017)
Δlog(Manufacturer Price)						0.887^{***}
						(0.084)
Observations	145,357	145,357	145,357	145,357	11,265	11,265
Retail Firms	110	110	110	110	110	110
Implied Tax-Inclusive Price Change	0.024	0.025	0.023	0.230	0.020	0.080
P-Value for Test of Constant Tax-Inclusive Price	0.000	0.000	0.000	0.000	0.270	0.000
			Placebo			
ΔPlacebo	-0.006^{*}	-0.004	0.001	-0.029	-0.016	-0.004
	(0.003)	(0.003)	(0.002)	(0.017)	(0.012)	(0.009)
Δlog(Manufacturer Price)						0.642^{***}
						(0.053)
Observations	253,123	253,123	253,123	253,123	11,534	11,534
Retail Firms	106	106	106	106	105	105
Bandwidth	6 weeks	6 weeks	6 weeks	6 weeks	6 weeks	6 weeks
MRS FE?	No	No	No	No	Yes	Yes
Inventory Lot FE?	No	Yes	Yes	Yes	No	No
Difference Length	1 week	1 week	2 weeks	1 week	1-4 weeks	1-4 weeks
Restricted to First Week Sales?	No	No	No	No	Yes	Yes

This table reports estimates of Equation (1) – other variables in that equation are included but not reported. An observation is an inventory-lot-week. The outcome is the log of the tax-exclusive price per gram charged by the retailer to consumers (except for in column (4) which is the same outcome, but not logged). MRS stands for manufacturer-retailer-strain fixed effects. The estimates are weighted by the total grams sold in the first week of the difference. The P-value tests the null hypothesis that the tax-inclusive price remained constant as predicted by TIV. For the placebo regressions, we repeat the analysis one year later. These regressions are estimated with reghdfe in Stata. In the last two columns we only include observations in their first week of being sold at retail and only if the cannabis was also purchased from the processor in that same week. Standard errors twoway-clustered by manufacturer and retailer are in parentheses (Cameron et al., 2011). *5% significance level. **1% significance level. ***0.1% significance level.

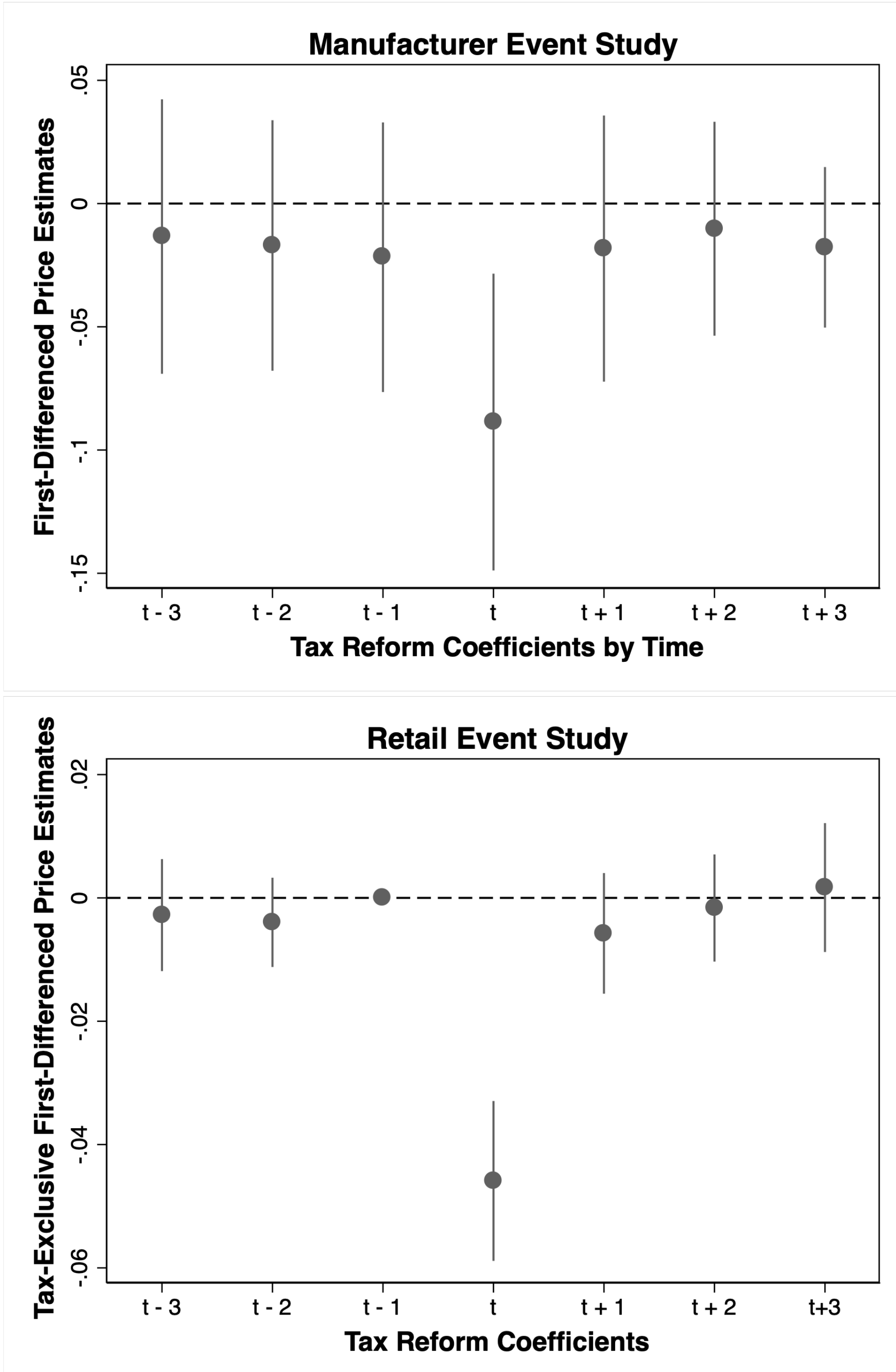
Figures

Figure 1: Histogram of Manufacturer Price Changes



This figure plots the dependent variable, $\Delta log(Price)$, for each retail-manufacturer-strain tuple in the baseline estimate sample. Weekly log manufacturer price changes are in the top panel and monthly log manufacturer price changes are in the bottom panel. The period of the tax reform is marked by the hollow green histogram and the other surrounding periods – six weeks or months pre- and post-reform – are marked by the gray histogram (the months affected by 4/20 are excluded from that histogram). The width of each bin is 0.04, so that all price changes within 2% of zero are included in the bin centered around zero. The dashed lines from right to left indicate: full pass-through predicted by models of firm behavior (-28.7%), our alternative benchmark of pass-through (-17.7%), and zero (i.e no price change). Outliers outside the interval [-1,1] are excluded.

Figure 2: Manufacturer and Retail Price Event Studies



This figure plots estimates of Table 2 Column (2) (top panel) and Table 3 Column (2) (bottom panel) with additional leads and lags of $\Delta TaxReform$. The plotted coefficients are leads and lags of $\Delta TaxReform$. We include in the regression (but do not plot) leads and lags are for periods $t - 4$ and before and $t + 4$ and after as is standard in event study designs. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 2 and 3 for regression details.

Appendices

A A model of tax pass-through in an imperfectly competitive supply chain

In this Appendix, we write a stylized model of the supply chain with imperfect competition in the style of Spengler (1950) and Weyl and Fabinger (2013) (whose notation we largely adopt) with the addition of percent-based taxes. In our model, manufacturers choose prices charged to retailers, who then charge prices to consumers. Both types of firms engage in imperfect competition characterized by a type-specific *conduct parameter*; Weyl and Fabinger (2013) show that this structure nests models of imperfect competition including homogeneous products oligopoly, differentiated Nash-in-prices, and monopolistic competition. We focus on the case in which firms within each layer are symmetric and on the unique symmetric equilibrium. We conclude this section by considering asymmetric imperfect competition.

Suppose there are n_r symmetric retailers, denoted by i, each producing a single product. Each faces constant marginal cost of p_1, the tax-inclusive price charged by manufacturers. Each firm sells quantity $q_i(p_i, \boldsymbol{p}_{-i})$ that depends on its own and others' prices. Products are (weak) substitutes: $\sum_{k \neq i} \frac{\partial q_k}{\partial p_i} / \frac{\partial q_i}{\partial p_i} \geq 0$ for all i and the profit function is concave. Market demand is smooth and symmetric-at-symmetric prices: $q(p_0) \equiv q_i(p_0, p_0, \cdots, p_0)$ for any i where p_0 is the tax-inclusive price paid by consumers. The price elasticity of demand is $\epsilon_D = -(dq/dp_0) \cdot (p_0/q)$ and is assumed to be greater than unity (i.e. if all retailers increased their prices infinitesimally, industry revenues would decrease). Retailers face a tax τ_r that is implemented as an excise tax: the unit net-of-tax revenue earned by the retailer is $p_0/(1+\tau_r)$.

Following Weyl and Fabinger (2013), we model competition by assuming the elasticity-and-tax adjusted Lerner index is set equal to a conduct parameter θ_r. That is,

$$\frac{p_0 - (1 + \tau_r)p_1}{p_0}\epsilon_D = \theta_r. \tag{2}$$

While we take θ_r to be exogenous with respect to the tax rate, we note that in general θ_r might depend on p_0 and therefore indirectly on τ_r. Given the assumptions above, given p_1 there exists some unique $p_0(p_1)$ that satisfies Equation (2). Weyl and Fabinger (2013) show that when products are weak substitutes, the conduct parameter ranges from 0 (perfect competition) to 1 (monopoly).

Suppose retailers purchase their products from n_m upstream manufacturers denoted by j. Manufacturers face a constant marginal cost mc. We again focus on the unique symmetric equilibrium. Manufacturers face effective symmetric-at-symmetric prices demand $\frac{n_r}{n_m}q(p_0(p_1))$. Their effective demand elasticity is $\epsilon_{D_m} \equiv \frac{dq}{dp_1}\frac{p_1}{q} = \frac{dq}{dp_0}\frac{dp_0}{dp_1}\frac{p_1}{q}$. Note that $\frac{dq}{dp_0} = \epsilon_D \frac{q}{p_0}$. Rearrange Equation (2) to write $p_0 = \frac{\epsilon_D}{\epsilon_D - \theta_r}p_1(1+\tau_r)$ so $\frac{dp_0}{dp_1} = \frac{\epsilon_D}{\epsilon_D - \theta_r}(1+\tau_r)$. Therefore, at the unique equilibrium,

$$\epsilon_{D_m} = (1+\tau_r)\frac{\epsilon_D}{\epsilon_D - \theta_r} \cdot p_1 \frac{\epsilon_D}{p_0}. \tag{3}$$

Manufacturers face a tax rate τ_m implemented as a gross receipts tax, so if the unit price paid by retailers to manufacturers is p_1, the unit net-of-tax revenue earned by manufacturers is $(1-\tau_m)p_1$. Though this does not match the retail side, we employ this definition to match our empirical setting; our results are invariant to this distinction. We again model imperfect competition through a conduct parameter: manufacturers choose prices such that

$$\frac{p_1 - mc/(1-\tau_m)}{p_1}\epsilon_{D_m} = \theta_m. \tag{4}$$

As above, given our assumptions, $\theta_m \in [0,1]$.

Equilibrium is characterized by the prices p_0 and p_1; Equations (2) and (4) are the equilibrium conditions. As ϵ_{D_m} is well-defined in equilibrium, the conditions can be solved to yield equilibrium relationships

$$p_0 = \frac{(1+\tau_r)\epsilon_D^2 mc}{(1-\tau_m)(\epsilon_D - \theta_m)(\epsilon_D - \theta_r)}, \quad p_1 = \frac{\epsilon_D mc}{(1-\tau_m)(\epsilon_D - \theta_m)}. \tag{5}$$

By substituting these expressions into Equation (3), we demonstrate that in equilibrium, $\epsilon_{D_m} = \epsilon_D$; τ_r does not change the effective demand elasticity for manufacturers, and so it does not enter their pricing condition.

To understand these relationships, it is useful to consider a perfectly competitive supply chain i.e. $\theta_r = \theta_m = 0$. In this case, manufacturers price equal to their effective marginal cost, $p_1 = \frac{mc}{1-\tau_m}$. Retailers face this price as their marginal cost and set $p_0 = mc\frac{1+\tau_r}{1-\tau_m}$. Now consider a monopoly-monopoly supply chain: $n_m = n_i = 1$ and $\theta_m = \theta_r = 1$. In this case, the equilibrium conditions reduce to a double marginalization condition. $\theta \in [0, 1]$, therefore, can be thought of as the extent to which firms behave as though they have monopoly power.

The taxing authority earns per-unit taxes of $rev(\cdot) = p_0\frac{\tau_r}{1+\tau_r} + p_1\tau_m$. Substituting Equation (5),

$$rev(\tau_r, \tau_m) = \frac{\epsilon_D mc}{(1-\tau_m)(\epsilon_D - \theta_m)}\left[\frac{\epsilon_D \tau_r}{\epsilon_D - \theta_r} + \tau_m\right]. \tag{6}$$

Define $p_0(\tau_r, \tau_m)$ as the tax-inclusive unit price faced by consumers as a function of taxes. Given per-unit revenue R, TIV holds if $p_0(\tau_r, \tau_m) = p$ for all $\{\tau_r, \tau_m \text{ s.t. } rev(\tau_r, \tau_m) = R\}$. To see that TIV holds in this model under perfect competition, define $\tau = \frac{1+\tau_r}{1-\tau_m}$. Then $\theta_r = \theta_m = 0$, $p_0 = mc{\cdot}\tau$, and $rev(\tau_r, \tau_m) = \frac{mc}{1-\tau_m}(\tau_r + \tau_m)$. Note that $\tau_r = \tau(1-\tau_m) - 1$. Substitute into $rev(\cdot)$ to write $rev(\tau) = mc(\tau - 1)$. In other words, given some τ, a policymaker can freely move one of τ_r or τ_m, solve for the other, and hold p_0 and revenue constant.[22]

Now consider the case when $\theta_r > 0$. If $\tau = \frac{1+\tau_r}{1-\tau_m}$, the term in brackets in Equation (6) cannot be simplified to a function of τ alone. Given some p, if a policy maker shifts τ_r and τ_m to hold $p_0 = p$, rev must change. Thus, TIV fails whenever $\theta_r > 0$. Intuitively, the percentage taxes act as demand shifters, but manufacturers do not internalize retailers' responses to retail percentage taxes because their effective demand elasticities are unchanged.

Although these results suggest that we should not expect TIV to hold empirically (as retailers are likely imperfectly competitive in our setting), this model may allow us to set

[22]Similar logic applies whenever $\theta_r = 0$ if ϵ_D is constant: $p_0 = \frac{\epsilon_D}{\epsilon_D - \theta_m} mc{\cdot}\tau$ and $rev = \frac{\epsilon_D}{\epsilon_D - \theta_m}\frac{mc}{1-\tau_m}(\tau_r + \tau_m)$ so since $\tau_r = \tau(1-\tau_m) - 1$, $rev = \frac{\epsilon_D}{\epsilon_D - \theta_m} mc(\tau - 1)$.

expectations for empirical movements in prices; if we find evidence conflicting with these expectations, we can reject the (very general) assumptions of the model. Suppose the policy $\omega_1 = \{\tau_r = 0, \tau_m = \tau\}$ is replaced with $\omega_2 = \{\frac{\tau}{1-\tau}, 0\}$. From Equation (5), it is clear that $p_0(\omega_1) = p_0(\omega_2)$ and thus the market quantity remains constant. Manufacturers pass-through their tax savings and earn identical per-unit profits. However, the retailer's per-unit profits decrease because the τ savings on the manufacturers' prices is more than offset by the new retail tax of $\frac{\tau}{1-\tau}$. By the same logic, $rev(\omega_2) > rev(\omega_1)$ as $\frac{\epsilon_D}{\epsilon_D - \theta_r} > 1$.[23] In other words, a policy change that moves taxes from manufacturers to retailers yet maintains constant retail prices increases total tax revenues.

Consider the policy $\omega_3 = \{\tau', 0\}$ where $\tau' = \frac{p_1(\omega_1)\tau}{p_0(\omega_1) - p_1(\omega_1)\tau}$ is "naive-revenue-neutral": it would raise the same total revenue *if* the tax-inclusive retail price p_0 was the same under ω_3 as it is under ω_1; this closely matches our empirical setting. In this case, since $\tau' < \frac{\tau}{1-\tau}$, $p_0(\omega_3) < p_0(\omega_1)$. Since $\epsilon_D > 1$, per-unit profits for both types of firms and per-unit tax revenues increase.

Finally, consider $\omega_4 = \{\tau'', 0\}$, where τ'' is chosen to be "true-unit-revenue-neutral": $rev(\omega_1) = rev(\omega_4)$. Since $rev(\omega_3) > rev(\omega_1)$, $\tau'' < \tau'$ and thus ω_4 increases per-unit profits for both retailers and manufacturers relative to profits under ω_1 beyond the increases realized under ω_3. These results are consistent with the notion that, under imperfect competition, ad valorem taxes improve welfare over unit taxes (Suits and Musgrave, 1953; Skeath and Trandel, 1994; Keen, 1998).

In summary, the combination of retail market power and percent taxes leads traditional TIV to fail. However, revenue-neutral policies (whether "naive" or "true") that shift taxes from manufacturers to retailers lead to full pass-through from manufacturers to retailers and a decrease in tax-inclusive retail prices faced by consumers, *regardless of the precise model of imperfect competition or demand*.[24]

[23]Note that since $p_0(\omega_1) = p_0(\omega_2), \epsilon_D(\omega_1) = \epsilon_D(\omega_2)$ by construction.

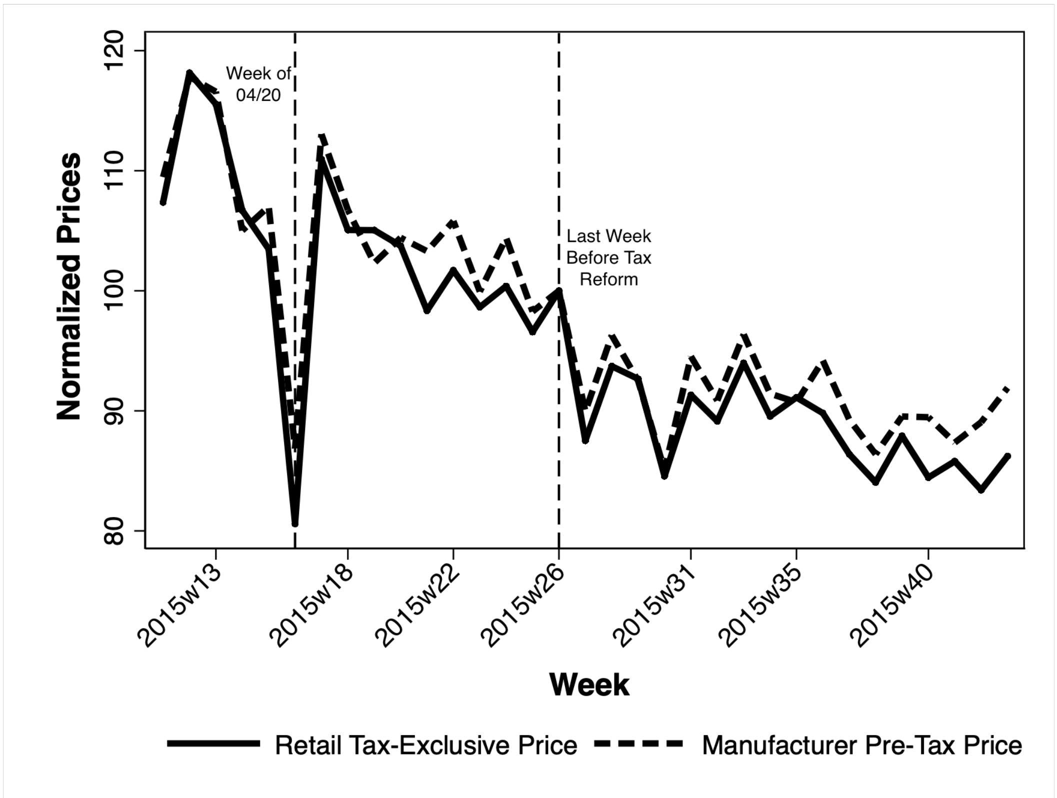
[24]Equation (5) shows that manufacturers should fully pass-through changes to tax rates whether those changes are revenue-neutral.

While we have assumed symmetry for ease of exposition, our results are robust to relaxing this assumption. Suppose manufacturers and retailers behave according to individual conduct parameters $\theta_{r,i}, \theta_{m,j} \in [0, 1]$. Weyl and Fabinger (2013) show that equilibrium can be characterized by taking the quantity-weighted mean of the conduct parameters as a market-level conduct parameter. If all of the firm-level conduct parameters are in the range $[0, 1]$, the market-level conduct parameters will also be within this range and thus the previous discussion follows.[25]

[25] Indeed, even if some firms were to behave (paraphrasing Weyl and Fabinger) "less competitively than monopolists" or if some products were effectively complements, the logic follows as long as the quantity-weighted average conduct parameter is less than ϵ_D.

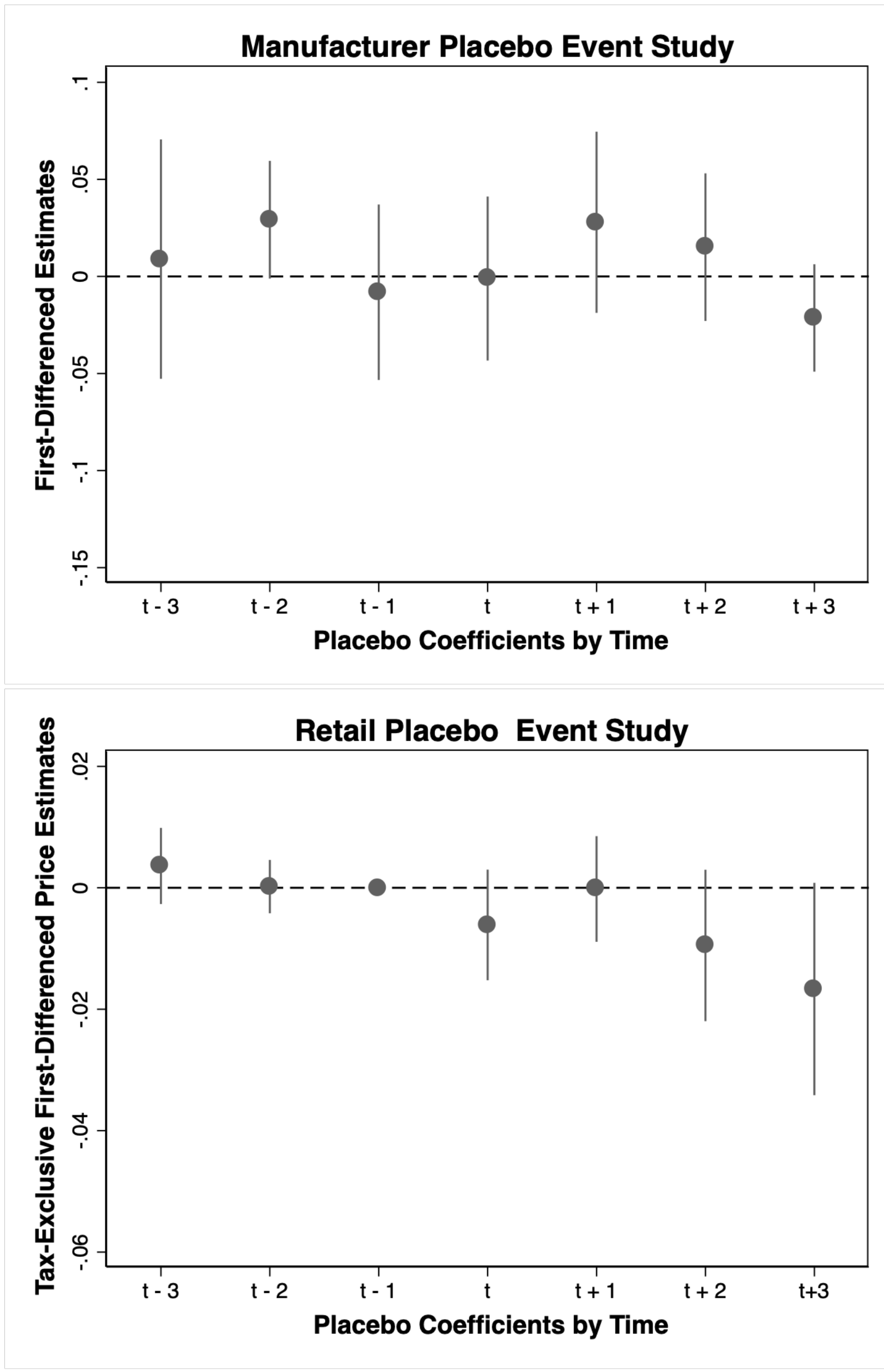
B Additional Figures

Figure B.1: Matched Retail and Manufacturer Prices



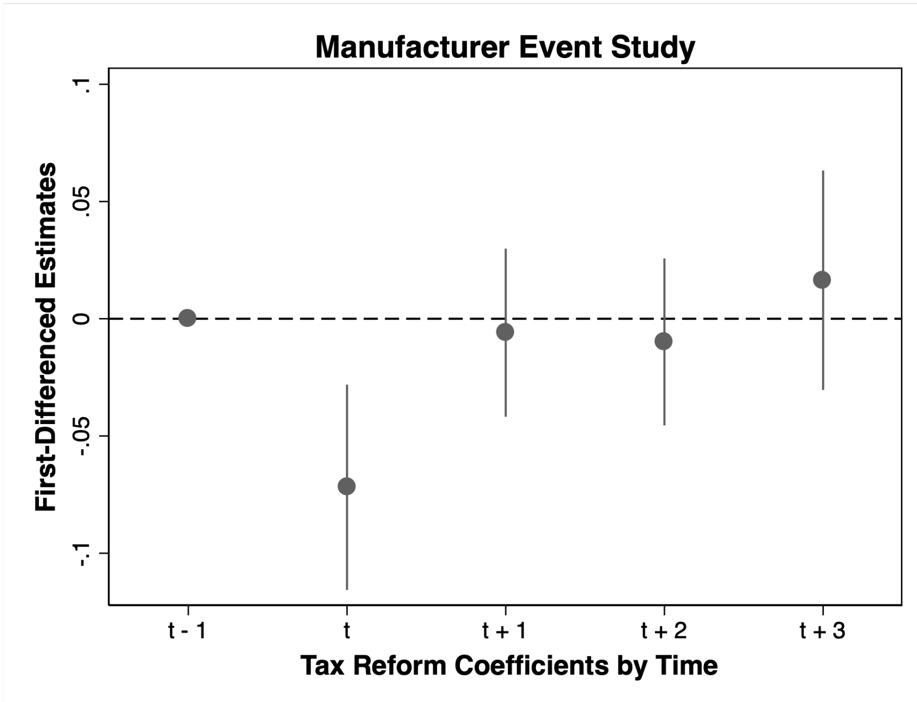
This figure plots raw average prices in Washington's cannabis industry for four months before and after the tax reform, normalized to 100 in the week before the reform. For each week, we take inventory lots in their first week of sale and match them with the price paid to the manufacturer, restricting observations to those for which the first retail sale and manufacturer sale both happened pre- or post-reform (before any applicable taxes are paid from the manufacturer to the government). This therefore illustrates the relation between retailer per-gram revenue and variable costs. The left dashed line in the figure marks 4/20 (an industry promotional period) and the right dashed line marks the week before the tax reform.

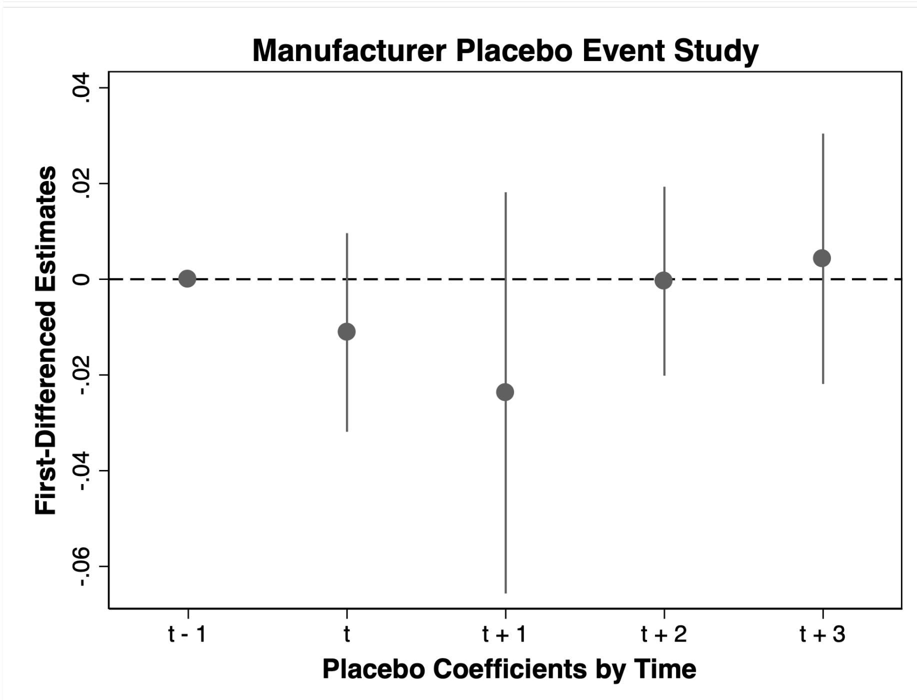
Figure B.2: Placebo Manufacturer and Retail Price Event Study



This figure repeats 2 one year later as a placebo robustness check. The plotted coefficients are leads and lags of $\Delta Placebo$. We include in the regression (but do not plot) leads and lags are for periods $t-4$ and before and $t+4$ and after as is standard in event study designs. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 2 and 3 for regression details.

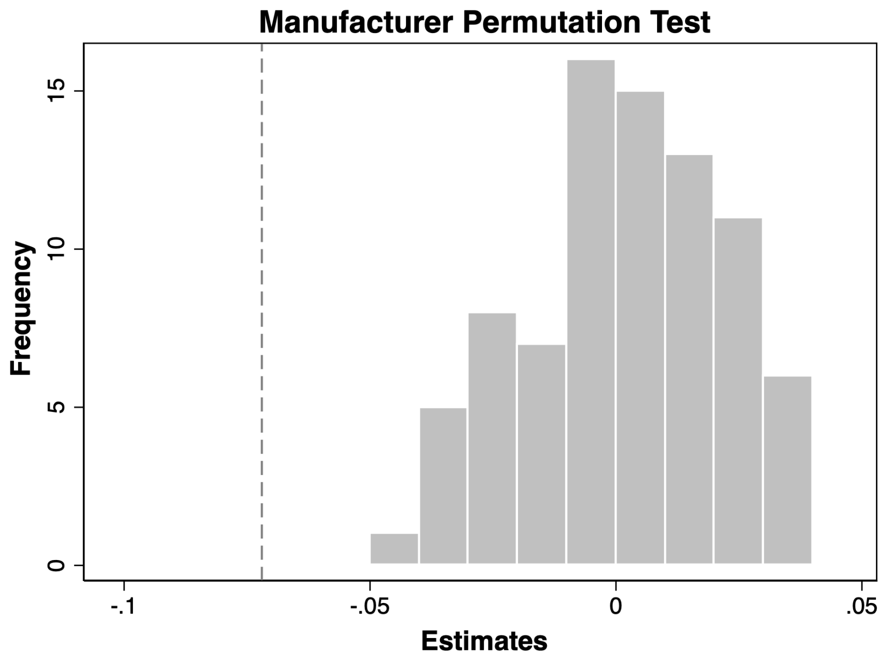
Figure B.3: Manufacturer Monthly Event Studies

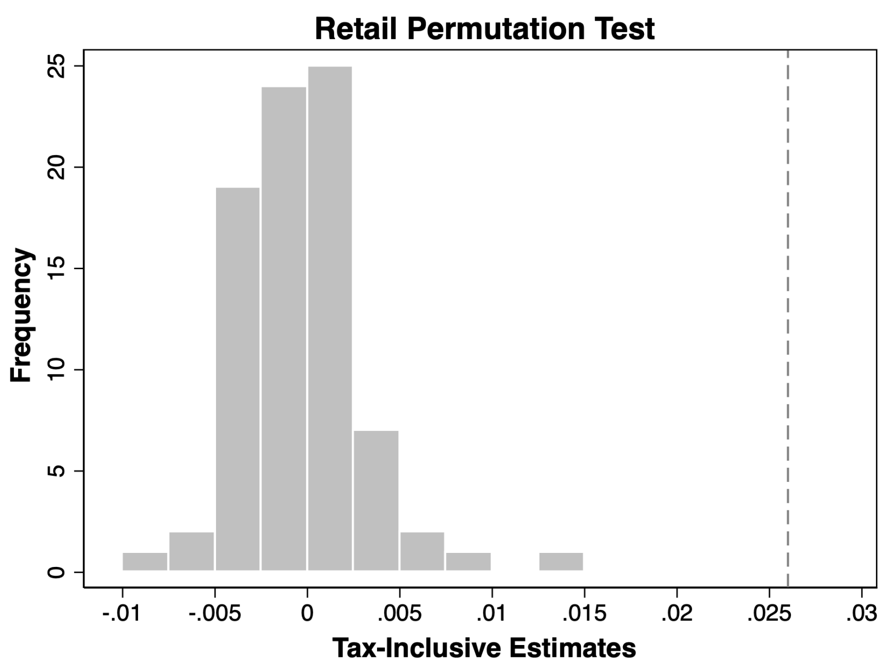




This figure repeats the top panels of Figure 2 and Figure B.2 for monthly, rather than weekly, differences. The plotted coefficients are leads and lags of $\Delta TaxReform$. The dots indicate the point estimates and the lines indicate 95% confidence intervals. We include in the regression (but do not plot) leads and lags are for periods $t-4$ and before, and $t+4$ and after as is standard in event study designs. We additionally exclude periods $t-2$, $t-3$ from our plot (even though they are included in the regressions) because first differences for both of those time periods include effects of the holiday 4/20 and are thus contaminated – we expect there to be a non-zero response in those months.

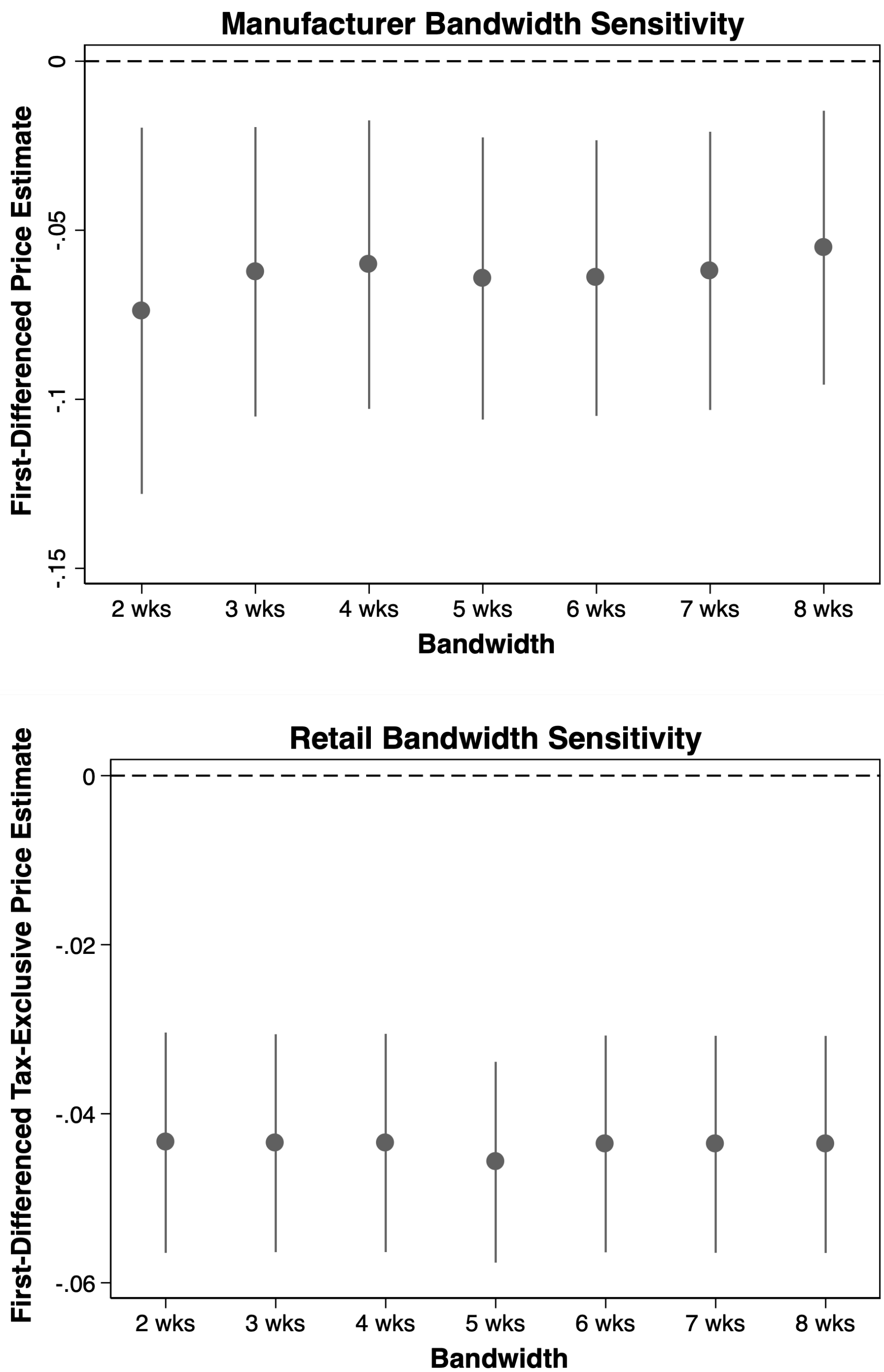
Figure B.4: Permutation Test Histograms





This figure conducts a placebo permutation test. The top panel is for our manufacturing price estimates and the bottom panel is for our retail tax-inclusive price estimates. The histogram plots placebo estimates using our baseline specification, but reassigning treatment to each week in our data, except those within two weeks of the tax reform, Black Friday, and 4/20. We also do not begin these estimates until there are at least 5,000 observations in our manufacturing regressions (which occurs in mid-March 2015). The gray dashed line marks our estimated effect in each figure. These lines are well outside the next most extreme estimates. There are 83 observations in each of our permutation tests, so the implied p-value is 0.012 (=1/83).

Figure B.5: Manufacturer and Retail Price Bandwidth Choices



This figure plots estimates of Table 2 Column (2) in the top panel and Table 3 Column (2) in the bottom panel, varying the bandwidth. The bandwidth in our baseline specifications is 6 weeks. The estimates plotted are for the coefficient on $TaxReform$. The dots indicate the point estimates and the lines indicate 95% confidence intervals. See the notes for Tables 2 and 3 for regression details.

C Data Cleaning Appendix

In this appendix we detail our data cleaning procedure. We begin by detailing our methods for cleaning prices in the face of changing reporting requirements and tax rates. We then discuss other cleaning steps to transform the raw data into the set used in our analyses.

The retail sales prices reported by firms in the "seed-to-sale" traceability system were supposed to be all-tax-inclusive pre-reform and tax-exclusive post-reform. However, compliance varied from firm to firm and changed over time. For example, some firms reported prices with the sales tax included and some reported prices without the sales tax.

This reporting confusion means that we must infer, for each firm, how they reported their prices and therefore the true tax-inclusive and tax-exclusive prices they charged. For each firm-week, we assign a "multiplier" that reveals the relationship between the reported price and the price faced by consumers. This chosen multiplier is selected from a set of multipliers based on possible tax rates for the firm. We merge in the state and local sales tax rates for each firm in order to construct this choice set.[26] To understand the relationship between the multiplier, reported prices, and faced prices, consider the following equation:

$$Price_{Consumer} = Price_{Reported} \times Multiplier$$

We algorithmically determine which tax-based multiplier makes the prices faces by consumer's ($Price_{Consumer}$) most round for each week, where roundness is the closeness of the price to a 25 cent increment of a dollar. For each product type, $Price_{Reported}$ is the modal observed price for the week, where idiosyncratic discounts have been removed.[27]

We consider two orthogonal methods of determine the proper set of multipliers. Our results are robust to the method used. Ultimately, we find the modal firm never included

[26]For five firms, the state and local tax rates do not match the rates they are using, so we adjust these. And a few firms do not ever change their local tax rate for reporting purposes—we make that adjustment as well. This transforms these firms from very unround to very round, but otherwise has approximately no effect on the data as the difference between the statutory and reported local tax rates is very small.

[27]We determine that a price is a one-off discount if the price for that transaction is 5% to 95% (in increments of 5 percentage points) or 33%/66.67% less than the previously reported price.

the sales tax, included the excise tax pre-reform, and excluded the excise tax post-reform.

Cash Market Identification In order to determine how each firm reports their prices in the traceability system, we take advantage of two characteristics of retail prices. First, publicly advertised prices (or 'list' prices) are nearly universally all tax-inclusive. Second, retailers nearly always choose to set prices in whole-dollar or (rarely) quarter-dollar increments.[28] We use these two facts to determine the difference between the list prices faced by consumers and the prices reported in the traceability system.

We assign each firm a multiplier before and after the tax change. We begin by assigning the modal firm's multiplier choices to all firms—all firms' prices were adjusted by the state and local sales tax pre-reform and all firm's prices were adjusted by the excise tax plus state and local sales tax post-reform. We then make the following adjustments based on the results from our algorithm:

1. We leave prices unadjusted (i.e. a multiplier of 1) where are algorithm finds that this choice maximizes roundedness and at least 85 percent of weekly sales are round with this multiplier choice.[29] This applies to 16% of firms.

2. We adjust the multiplier post reform to account for only the excise tax when the algorithm finds that this choice maximizes roundness and at least 85 percent of weekly sales are round with this multiplier choice. This applies to one firm (out of 110).

   There are three additional firms for whom an only excise tax adjustment makes them most round, but their roundness in the immediate post period is less than 85 percent. We leave two of the firms alone because they were also left alone in the pre-reform period because of unroundness and we could either adjust them both before and after

[28]We verified this through conversations with retailers as well as using historical menus available through The Internet Archive and a full set of menus for almost all firms we took screen shots of on 7/18/2017.

[29]For the 4.5% of firms that suggest the multiplier could be 1 but are quite unround, there is too much uncertainty to confidently make an adjustment. Leaving these firms' multipliers unchanged, if wrong, will bias our estimates towards our main null hypothesis in the retail section of the paper—that firms did not adjust their prices in response to the reform.

the reform or leave them both alone with similar effects to the log price change. The third firm becomes more round a few weeks after the reform and keeps this multiplier through the end of our data (and we have confirmed the multiplier in the menu screen shots), so we make this multiplier adjustment.

3. There are two firms for whom the multiplier that makes them round post-reform is the excise tax + state and local sales taxes divided by the state and local sales tax rate. In both cases, we have clear evidence that this is because they adjusted their prices post-reform by making their prices sales-tax exclusive post-reform. One firm keeps this choice permanently and we see this in the menu screen shots at the end of our data. The other firm eventually adjusts to the modal firms' multiplier. Our assumption keeps prices roughly constant through this reporting change.

Product Batch Price Stability To provide additional evidence that our multiplier decisions are not systematically biasing our estimates, we consider a completely different mechanism for determining multipliers—we pick the set of multipliers that makes the tax-inclusive prices for the most number of inventory lots for a given firm the same pre- and post-reform.

There are a couple of reasons why this is a reasonable alternative to consider. A number of inventory lots did leave prices constant in response to the tax reform and the main null hypothesis in our retail analysis is that firms did not change their tax-inclusive prices—this is what we would expect if the tax reform was indeed tax invariant.

We consider two variants of this. One is to begin with the modal firms' multipliers and adjust it to another multiplier if it decreases the number of price changes by any margin. The second variant is to begin with our estimates based on roundedness and then make adjustments for firms that under the best set of multipliers leaves at least 25% of their inventory lots constant in response to the reform. The latter changes the multipliers for only four firms and three of those four leave the percent price changes quite similar. The former method decreases our baseline estimate by 0.4 percentage points and the latter decreases our baseline estimate by 0.2 percentage points. This evidence strongly supports our price cleaning methods and suggests that any remaining bias is extremely small.

Additional Cleaning

In addition to adjusting retail prices, we also drop some extreme outliers in the data. In particular, we drop all wholesale transactions with a usable weight above 2,500 grams[30] and all retail transactions if the usable weight was above 28.5 grams.[31] We also drop all wholesale or retail price per grams above $42.[32] We censor the THC content data if it is zero or above 40 in both the manufacturer and retailer data.[33] We also drop wholesale prices less than $1. This effectively drops samples from our data, which are sold well below market value. We typically see these as the first recorded sale from a parentlot.

Lastly, we drop some firms or firm-days in our data set. In particular, we require for each firm that the first sales transaction occurs two months before the tax reform and continues to have transactions through the two months after the reform (either because they had not yet opened, had closed, or because they took a long hiatus from selling any cannabis). A few retailers conduct a "soft opening" by opening briefly, closing for more than a month, and then re-open permanently. In these cases, we drop the first brief selling period and consider their first activity date the first date upon re-opening in our data. We also drop 20 retail firms for whom at some point in the 8 weeks before or after the reform report their data only once per day—this is a clear indicator of poor overall data quality and, because of this, determining the appropriate multipliers for these firms is difficult.

[30] This is about 0.025% of wholesale transactions.

[31] This is because the maximum legal sale was one ounce. This step drops 0.15% of retail transactions.

[32] This is less than 0.03% of wholesale transactions and less than 0.04% of retail transactions.

[33] This affects 0.2% of wholesale transactions and 5% of retail transactions.