Revisiting the Effects of Cigarette Taxation on Smoking Outcomes*

Vinish Shrestha†

26 June, 2024

Abstract

This study reassesses the efficacy of cigarette taxation in curtailing smoking by leveraging recent advancements in the difference-in-differences (DiD) literature to account for heterogeneous treatment effects. Using data from the Behavioral Risk Factor Surveillance System Selected Metropolitan/Micropolitan Area Risk Trend (BRFSS SMART) for the sample periods 2004-2010 and 2015-2020, the study reveals three key findings. Firstly, the TWFE estimate for the 2004-2010 sample is only 48% of the average treatment effect on the treated (ATT) estimate obtained through the DiD framework. Secondly, event-study type estimates demonstrate a gradual increase in magnitude following the treatment year, highlighting dynamic treatment effects overlooked by the TWFE estimate. Third, the ATT estimate for the 2015-2020 sample is approximately 66% of the ATT estimate for the 2004-2010 sample. Overall, the study underscores the potential bias towards zero in elasticity estimates when relying solely on TWFE models.

Keywords— Cigarette taxation, Difference-in-differences, Treatment heterogeneity, Dynamic treatment effects, Elasticity

JEL Codes— I10, I18, D00, B23, H20

Statements and Declarations: No funding was acquired for this research. There is no conflict of interest.

*I would like to thank Catherine Maclean, Erik Nesson, and the participants of the Tobacco Online Policy Seminar (TOPS) series as well the 17th International WEAI conference for their valuable comments and suggestions. All errors are my own.

†email vshrestha@towson.edu, Towson University, Economics Department, Maryland, USA.

1 Introduction

Cigarette taxation serves as a widely used policy instrument to both diminish smoking and bolster revenue in the United States. To evaluate the efficacy of higher cigarette taxes (prices) on smoking outcomes, researchers have increasingly relied on the two-way fixed effect (TWFE) specifications.[1] These studies use continuous measure of cigarette taxes (prices) and utilize within state variation (mostly increases) in cigarette taxes (prices) over the years in a multiple-treatment and multiple-control group framework to identify the target parameter.

This study revisits the literature evaluating the effects of cigarette taxes on smoking outcomes to provide a more accurate measure of the effect of the impact of cigarette taxation in reducing smoking. Recent advancements in staggered difference-in-differences literature have highlighted concerns regarding the TWFE estimator (De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Goodman-Bacon (2021)), raising concerns regarding the accuracy of previous estimations. The TWFE models pose restrictions of homogeneous effects across group and the length of exposure to the treatment, potentially lending to biased estimations when effects are not uniform. A critical issue highlighted is that of the negative weighting problem, where varying treatment effects with exposure length make early treated groups inappropriate comparators for units treated later in the sample period. This problem is particularly acute when a significant portion of units eventually receive treatment. Given that 38 states increased cigarette taxes at least once between 2004 and 2010, this issue holds significant relevance when using TWFE models to evaluate the effects of cigarette taxes, especially considering the heterogenenous nature of these effects. By addressing these concerns, this study aims to shed new light on the impacts of cigarette taxation on smoking behavior, offering insights crucial for effective policy-making.

The study employs newer estimation techniques based on staggered difference-in-differences for multiple-group and multiple-time treatment framework that are less restrictive compared to the TWFE estimator to estimate the average treatment effect on the treated estimates ($A\hat{T}T$). The analysis utilizes a balanced panel data from the Behavioral Risk Factor Survelliance System Selected

[1] See Adda and Cornaglia (2006), Tauras et al. (2007), Carpenter and Cook (2008), Callison and Kaestner (2014), DeCicca et al. (2008), Harding, Leibtag, and Lovenheim (2012), C. Cotti, Nesson, and Tefft (2016), Hansen, Sabia, and Rees (2017), C. Cotti, Nesson, and Tefft (2018).

Metropolitan/Micropolitan Area Risk Trends (BRFSS SMART). The analyses are conducted for the 2004-2010 (early) and 2015-2020 (late) samples, respectively.[2] This allows for an inspection of cigarette taxes as a policy tool to curtail smoking in more recent years as well as the comparison of estimates across time.

The analysis begins with diagnostic checks and simulation exercises to understand the performance of TWFE models given the actual variation in cigarette taxes. By defining treatment as an indicator for cigarette tax increase (binary treatment), the findings from simulation exercises demonstrate that the TWFE estimates are severely biased towards zero for the 2004-2010 sample in situations of heterogenenous treatment effects by groups or relative time from the treatment period. This is because units treated in 2005 and 2006 pose negative weights when used as comparison units in 2008-2010 and 2010 calendar years/year, respectively.

Next, the analysis estimates the impacts of tax incidence using TWFE models and compares the estimates to $\hat{ATT}$ obtained from Callaway and Sant'Anna (2021) (CS from hereon) estimator, which is based on the concept of group-time treatment effects and is robust to problems associated with the TWFE estimator. Additionally, the TWFE estimates are compared with the results from: *i*) canonical event-study design, *ii*) interaction-weighted estimator proposed by Sun and Abraham (2021) (SA from hereon), and *iii*) event-study type design following Callaway and Sant'Anna (2021).

The findings from different approaches mentioned in the aforementioned paragraph demostrate effectiveness of tax incidence in improving smoking-related outcomes. However, the TWFE estimates are lower in magnitude compared to the overall $\hat{ATT}$ obtained from from CS estimator. This difference is considerably higher in the earlier sample period (2004-2010) for which the size of the TWFE estimate is only about 56% of the overall $\hat{ATT}$ from CS estimator. In other words, while the overall $\hat{ATT}$ point estimate from Callaway and Sant'Anna (2021) suggests that over a quarter of the reduction in prevalence of current smoking between 2004-2010 is attributed to cigarette tax incidence, the TWFE estimate only accounts for 14% of the reduction. The results from decomposition following Goodman-Bacon (2021) point out that the magnitude of the TWFE estimate is suppressed towards zero since a huge weight (31%) is placed on cases that use units treated early on in the sample as comparison for those treated later in the sample. The problem of

[2] The in between years are not included due to lack of BRFSS SMART data for these years.

bad comparison is not as dire in the case of the later sample (2015-2020) since the majority of units remain untreated throughout the sample. Moreover, the overall $A\hat{T}T$ magnitude for 2015-2010 sample is only 66% of the $A\hat{T}T$ in 2004-2010 sample. The diminished impacts of cigarette taxes observed in the recent sample periods could potentially be attributed to a reduced elasticity among smokers who continued to smoke despite early tax increases.

When estimating the dynamic effects of cigarette tax incidence, the findings from *i*) canonical event-study approach, *ii*) SA approach, and *iii*) event-study type method from CS all demonstrate gradual but increasing effects of tax incidence by the length of exposure to treatment. A simple comparison between the TWFE estimate and the post treatment event-study estimates show that the TWFE estimate only amounts to 46% of the average post-treatment estimates in the 2004-2010 sample. While there is a noticeable trend-break in estimates following the treatment period for the later sample (2015-2020), the majority of these estimates are statistically insignificant. Therefore, it is important to interpret these findings with caution. Moreover, the pre-treatment estimates in the event-study design pertaining to both the early and later samples and using different approaches fail to show any evidence of systematic differences in smoking outcomes between the treated and untreated units prior to the treatment. This is in support of the identification assumption used in the study.

This study's findings carry significant weight for policy-makers, shedding light on the potential underestimation of the impacts of cigarette taxation on smoking outcomes, particularly during years between 2004 and 2010. The repercussions of these underestimated effects extend to elasticity estimates, which form the backbone of welfare analysis in evaluating the costs and benefits of tobacco control policies. Moreover, given the increasing interest among practitioners in evaluating the multifaceted effects of cigarette taxes on outcomes beyond smoking, such as birth and health outcomes (Markowitz (2008), Markowitz et al. (2013), Patrick et al. (2016), Hoehn-Velasco, Pesko, and Phillips (2023), Friedson et al. (2023)), alcohol and marijuana consumption (Decker and Schwartz (2000), Farrelly et al. (2001), Picone, Sloan, and Trogdon (2004), Williams et al. (2004), Shrestha (2018), Anderson, Matsuzawa, and Sabia (2020)), and body mass index (BMI) (Gruber and Frakes (2006), Baum (2009), Conway and Niles (2017)), the estimates provided here serve as crucial "First Stage" effects in understanding the wider implications of these policies. Furthermore, accurate estimations of cigarette taxation can assist in the effective distribution of resources by governments and public health agencies for the administration of tobacco control programs and initiatives. As such, there is a clear need for more precise and accurate estimates of the effectiveness of cigarette taxes in reducing smoking rates to guide informed and impactful policy decisions.

A Brief Summary of the Literature. The empirical methods used to evaluate the impacts of tobacco regulations can be broadly grouped into two segments: pre and post 2000 studies. Chaloupka and Warner (2000) provide a review based on the pre-2000 studies and report that most of the price elasticity estimates of cigarettes range from -0.3 to -0.5.[3] However, the review majorly relies on the studies that use aggregated time series or cross-sectional data, which makes estimation more suseptible to omitted variable bias (Lewit and Coate (1982), Mullahy (1985), Jones (1989), Wasserman et al. (1991), Seldon and Boyd (1991), Sung, Hu, and Keeler (1994), Barnett, Keeler, and Hu (1995)). For instance, geographic units that increase cigarette taxes may have different socio-demographic characteristics as well as higher prevalence of anti-smoking sentiments even after controlling for income and other tobacco control policies, which may affect both the passage of higher cigarette taxes as well as smoking outcomes.[4]

The post-2000 studies to a large extent use individual level data and are methodologically based on TWFE models. These studies utilize some combination of spatial and temporal variation in cigarette taxes to identify the effects on smoking by controlling for both geographical unit (state) fixed effects and time fixed effects. While geographical unit fixed effects account for the time invariant unobservables, time trends in the model absorb common trends over time across all geographical units. Given that there exists a sufficient within variation in cigarette taxes across units, theoretically the post-2000 studies should improve upon the pre-2000 studies as unobservables such as anti-smoking sentiments are accounted by the unit fixed effects as long as the sentiments

[3]The consensus provided by the Chaloupka and Warner (2000) study is that the range of elasticity estimates are about the same for both extensive (smoking participation) and intensive margin (conditional demand).

[4]The pre-2000 studies have limited ability to control for such omitted factors. Moreover, from an empirical standpoint these studies are conducted during the time period with relatively meager increases in cigarette taxes compared to the post-2000 period. Other additional problems with many of the pre-2000 include simultaneity between cigarette prices and aggregated cigarette demand and usage of annual state level tax receipts on cigarette sales as the dependent variable. We refer the reader to the Chaloupka and Warner (2000) study for more detail regarding pre-2000 studies.

are time invariant during the span of the study. The majority of the post-2000 studies find that encouraging trends in smoking outcomes can be attributed to increases in cigarette taxes (prices) (Adda and Cornaglia (2006), Tauras et al. (2007), Carpenter and Cook (2008), DeCicca et al. (2008), Harding, Leibtag, and Lovenheim (2012), C. Cotti, Nesson, and Tefft (2016), Nesson (2017), Bishop (2018), C. Cotti, Nesson, and Tefft (2018), Pesko, Courtemanche, and Maclean (2020)), with a handful of studies showing small or non-exsistent effect (Callison and Kaestner (2014), Hansen, Sabia, and Rees (2017)).

Despite the extensive body of research on the effectiveness of tobacco control policies, a notable gap in the literature has emerged due to recent methodological advancements that address concerns related to TWFE models. In a recent comprehensive review highlighting the significance of tobacco regulations, DeCicca, Kenkel, and Lovenheim (2020) underscores the importance of exploring methodologies that are resilient to the challenges associated with TWFE models. The authors explicitly state, "*this is an important issue for the analysis of cigarette taxes that has not been sufficiently explored by researchers*."

This study contributes significantly to bridging the gap in the literature concerning the impact of cigarette taxation on smoking behavior. Firstly, it offers insights for policymakers and research practitioners by highlighting the limitations of relying solely on the TWFE model to estimate the effects of cigarette taxation. Through diagnostic checks, the study demonstrates the shortcomings of the TWFE model, particularly in the 2004-2010 sample, and proposes an alternative approach using recent methodological advancements such as the estimator proposed by Callaway and Sant'Anna (2021) to address issues like the "bad comparison" problem. Secondly, the study addresses a crucial aspect often overlooked in prior research: the parallel trend assumption.[5] By thoroughly investigating trends in smoking outcomes between treated and untreated groups prior to the treatment period using canonical event-study methods and more robust approaches proposed by Sun and Abraham (2021) and Callaway and Sant'Anna (2021), it enhances the credibility of its findings. Lastly, the study provides updated estimates of the relationship between cigarette taxation and smoking outcomes, shedding light on the evolving effectiveness of cigarette taxes over time.

[5] To our knowledge only C. D. Cotti et al. (2020) explores the parallel trend assumption. However, there are major differences between this study and the C. D. Cotti et al. (2020) study, whose main focus is to investigate whether cigarettes and e-cigarettes are economic substitutes.

While demonstrating the efficacy of cigarette taxes in earlier years, the findings suggest a potential decline in their impact on reducing smoking rates in recent years, thus informing policy decisions with more current and nuanced insights.

The study is structured as follows: Section 2 presents the data, while Section 3 outlines the methodologies employed. Section 4 examines the findings from diagnostic checks, Section 5 discusses the results, and Section 6 evaluates elasticity estimates. Finally, Section 7 concludes the study.

2 Data

2.1 BRFSS SMART

The primary data for smoking outcomes comes from the Behavioral Risk Factor Survelliance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART) for the years 2004-2010 and 2015-2020. The intervening years are omitted due to lack of SMART data. BRFSS data are typically used to construct state level estimates, but the SMART project was initiated to produce estimates for local areas defined as Metropolitan and Micropolitan Statistical Areas (MMSAs), delineated by the Office of Management and Budget (OMB).[6] The respondents are associated with MMSAs by their county of residence as reported during the survey. The MMSAs are represented by the Core Based Statistical Area (CBSA) codes that falls within the Federal Information Processing Standards (FIPS).[7]

The eligibility criteria to determine whether a particular MMSA is included in BRFSS SMART data depends on the number of observations in each weighting class. The weighting classes are based on age, race, and gender, which gives a total of 24 weighting classes.[8] MMSAs with weighting classes comprising less than 19 observations are excluded from the SMART dataset. The SMART data allows comparison of prevalence estimates across MMSAs as the same weighting criteria are

[6]According to the Centers for Disease Control and Prevention, the MMSA is defined as "a core area containing a substantial population nucleus, together with adjacent communities and all having a high degree of economic and social integration."

[7]There were a total of 944 MMSAs excluding Puerto Rico in 2010. The list of CBSA category (metropolitan vs. micropolitan), CBSA code, and CBSA title can be found on https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cls_cbsa/cbsa_countyassoc.htm.

[8]Some states do not use race in post-stratification. For these states only age and gender are used to form weighing classes, which gives 12 weighting classes.

used for all MMSAs. Each MMSA includes at least 500 individuals.

The number of MMSAs vary across the BRFSS SMART survey years as MMSAs can enter or exit the survey. For example, the 2004 survey includes 134 MMSAs, while the 2010 survey includes 198 MMSAs. The study focuses on the status of current smoker as the main dependent variable and create a balanced panel of the outcome variable collapsed at the MMSA-year level to assure that the findings are not driven due to differences in composition of MMSAs in the sample.[9] The BRFSS SMART sample weights are used during this process. As such, the 2004-2010 and 2015-2020 balanced panels include 108 and 95 MMSAs, respectively.

The MMSAs included in the balanced dataset for the BRFSS SMART survey years 2004-2010 and 2015-2020 are portrayed on the map using red color in panels A and B of Figure 1, while the green polygons represent MMSAs not covered by the survey. At least one MMSA is included in the balanced panel data for 46 states and there are more than two MMSAs in many states. The states that are not represented in BRFSS SMART (balanced panel) include Alaska, Hawai, North Dakota, and Rhode Island.

2.2 Cigarette taxes and tobacco control policies

The state level cigarette taxes are extracted from the consolidated files of Tax Burden of Tobacco for years 1970-2019 prepared by Orzechowski and Walker. This version of the file is obtained from the Centers for Disease Control and Prevention.[10] Using the reported excise tax and implementation date for tax changes, a binary variable is created to indicate whether a state increased cigarette taxes in a given year. The indicator variable takes the value 1 in the year of tax increase and the state retains this value for the rest of the panel, while years prior to the tax change is indicated by the value 0. Hence, the treatment assignment takes a value of 0 and 1 similar to the canonical difference-in-differences framework and maintains a staggered design.[11]

A handful of states experienced multiple tax increments within the span of the survey. For instance, Pennsylvania increased state cigarette taxes in July 2004 and November 2009. As both

[9] For this study, current smokers are defined as individuals who smoke cigarettes "every day" or "some days".

[10] See https://chronicdata.cdc.gov/Policy/The-Tax-Burden-on-Tobacco-1970-2019/7nwe-3aj9

[11] Once the treatment indicator turns on, it is not switched off. In fact, none of the states during the sample periods used in the study reduced cigarette tax after increasing it.

episodes of tax changes fall within the survey years 2004-2010, I use the earliest episode of tax change to denote the treatment assignment. In other words, the coding of treatment assignment is dependent on the earliest tax change date in cases of states with multiple tax changes.

To address the influence of smoke free laws, the analyses include controls for smoking ban in bars, which prohibits smoking tobacco products within the premises of bars or establishments where alcoholic beverages are served. These bans are typically enacted at the local or state level. Data on localities that implemented smoking ban in bars, along with the dates of implementation, are sourced from the American Nonsmokers' Rights Foundation (ANRF).[12] Core-Based Statistical Area (CBSA) population estimates for the years 2000 and 2010, obtained from the Census Bureau, are used to calculate the percentage of the population living under the bar ban policy for the survey years 2004-2010 and 2015-2020. Figure 11 exhibits the evolution of the percentage of population living under the ban ban policy from the year 2000 onwards in the United States. As shown, the percentage of population covered by the bar ban policy has increased from close to zero in early 2000s to almost 100% by the year 2020.

2.3 Other variables

Also, the study uses locality specific unemployment rate, the measure of anti-smoking sentiment in 1998-1999, changes in the proportion of current smokers between 1998-1999 and 2001-2002, and the log of population as covariates. The locality specific unemployment rate is calculated using data from the Current Population Survey (CPS) - Merged Outgoing Rotation Group Earnings Data for years 2000 and 2010. The unemployment estimates for 2000 and 2010 are merged with survey years 2004-2010 and 2015-2020, respectively. The data for variables used to form the anti-smoking sentiment measure comes from the Current Population Survey (CPS) Supplements: Tobacco Use for years 1998-1999. The anti-smoking measure is constructed using a principal component analysis in spirit of DeCicca et al. (2008) but the measure is aggregated at the locality level instead of the state level. Not all of the MMSAs listed in the balanced BRFSS SMART data are contained in CPS tobacco supplement. In the spirit of using all observations, I use a regression based imputation

[12] The list of local areas enacting smoking bans in public places (bars, restaurants, non-hospitality workplaces) are listed here.

method to impute values for the missing observations.[13] The changes in the proportion of current smokers between 1998-1999 and 2001-2002 are constructed by using data from the CPS tobacco supplement. The sub-figures in the Appendix section, Figure 10, show the negative relationship between the measure of anti-smoking sentiment and smoking-related variables, i.e., the proportion of current smokers and those who ever smoked cigarettes in 1998.

2.4 Descriptive Results

As shown in Table 1, thirty eights states increased cigarette taxes at least once between 2004-2010, while 18 states increased taxes between 2015-2020. The table suggests that cigarette tax hike (a binary treatment measure), is not systematically confined to a specific geographic region (at least within the sample periods of the study).The table also shows heterogeneity in the levels of tax increases. The average tax increase for the 2004 group (those increased cigarette tax in 2004) was only 34 cents, whereas the highest tax increase in the 2004-2010 sample is just over a dollar for the 2008 group. In the later sample, California increased cigarette taxes by $2 in 2017, while average tax increase for other years are below a dollar.

Table 2 provides descriptive statistics of some important variables separated by states that increased cigarette taxes versus states that did not. The first three variables portray the summary of smoking behaviors among individuals living in metropolitan/micropolitan areas (MMSAs) included in the BRFSS SMART sample. On average smoking-related variables between units exposed to tax increases versus those not exposed to tax increases are fairly similar across both sample periods. For example, around 20 percent of the 2004-2010 sample are current smokers and the magnitudes are similar across areas with and without tax increases. However, the state-specific per capita cigarette sales in the pre-treatment periods are higher on average among units exposed to tax changes in 2004-2010 sample. For example, the per capita cigarette sales in 1990 amount to 107 and 100 packs in states encompassing treated vs. untreated units, respectively. But this difference

[13] First, I regress anti-smoking measure on the log of cigarette sales in 2000, wage index, log of MSA population in 2000 and the state-level anti-smoking measure in 1998 using the MMSA level data with non-missing values for anti-smoking sentiment measure. Next, the coefficients are used to predict anti-smoking sentiment at the MMSA level and the non-missing values are replaced by the predicted values.

vanishes when focusing on the per capita cigarette sales in 2010.[14] Furthermore, as depicted in Table 5 in the Appendix, among states that increased cigarette taxes between years 2004-2010, tax hikes exhibit a positive association with levels of anti-smoking sentiments, while showing a negative association with cigarette sales prior to the hike.[15]

3 Method

We first discuss the TWFE model and briefly summarize some problems associated with TWFE in the context of cigarette taxation. Next, we thoroughly discuss the CS (Callaway and Sant'Anna (2021)) estimator. Throughout the analysis we utilize the variation in cigarette taxes in the two sample periods: *i*) 2004-2010; and *ii*) 2015-2020.

3.1 Two way fixed effect (TWFE) model

The TWFE specification is given by:

$$Y_{ist} = \alpha + \beta D_{st} + \sigma_i + \kappa_t + \eta BarBan_{ist} + \gamma X_i \times \kappa_t + \epsilon_{it} \tag{1}$$

where, Y_{ist} represents smoking outcome (i.e., the percent of current smokers) collapsed at the locality-level (metropolitan or micropolitan area) i within a state s at time t. D_{st} is an indicator variable representing the treatment assignment of an increase in cigarette taxes. If a state experiences multiple increases in cigarette taxes within the sample period, the earliest increase in cigarette taxes is used to define the treatment.[16] σ_i is the locality level fixed effects and captures the time invariant unobserved heterogeneity across localities, while κ_t absorbs common time trends

[14]Although the proportion of current smokers decreased by about 5 percentage points between 2004-2010 and 2015-2020 sample, the magnitudes are fairly similar across units with and without tax changes. While the average nominal cigarette tax is about 65 cents higher among the treated units in the 2004-2010 sample, the difference in tax amount is only 11 cents between the treated and untreated units in the later sample. Also, states with tax changes on average have a higher magnitude of anti-smoking measure and cover a greater percentage of people living under the bar ban policy in the earlier sample period.

[15]A detailed discussion is provided in Appendix section 02.B.

[16]For example, if a state increased cigarette taxes in both 2004 and 2009, the treatment turns on in 2004 and remains on for the following years.

in smoking outcomes.

The parsimonious treatment specification only includes the treatment indicator plus the area and time fixed effects. Additional specifications add the percent of a locality's population living under the bar ban at time t denoted by $BarBan_{ist}$ in equation 1. We defer from using time varying controls as much as possible due to the potential of post-treatment bias (see Rosenbaum (1984)). X_i is a vector of locality specific time invariant pre-treatment variables (the log of population, percent unemployed, a measure of anti-smoking sentiment in 1998-1999, change in the proportion of current smokers between 1998-1999 and 2001-2002) interacted with the year fixed effects.[17] However, controlling for covariates in a regression model as in equation 1 may not yield consistent estimates of ATT if treatment effects are heterogeneous across covariates, even when the conditional parallel trend assumption is satisfied; see Meyer (1995) and Roth et al. (2022) for discussion. The standard errors are clustered at the state level for all specifications.

Using the binary format of treatment is different from the majority of studies in the literature that use a continuous measure of cigarette taxes. This study uses the binary measure of treatment for two reasons. First, the binary measure complies with the staggered design of multiple-group and multiple-period difference-in-differences framework as discussed in the recent studies providing methodological advancements in the difference-in-differences literature (De Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021), Callaway and Sant'Anna (2021)). Second, the binary treatment variable maps the analysis to the more standard treatment effect literature including the difference-in-differences approach.

Next, to allow for the dynamic treatment we estimate a canonical event-study design. The event-study specification is given as below.

[17] The local area population and the percent unemployed is measured in years 2000 and 2010 for the BRFSS SMART sample years 2005-2010 and 2015-2020, respectively. The anti-smoking variable is constructed similar to DeCicca et al. (2008) study and is measured at the locality level in 1998. To construct the proportion of change in current smokers between 1998 and 2001 (years prior to the treatment), we rely on the tobacco supplement files from the Current Population Survey. The interaction of these pre-treatment variables with year indicators allow the effects of the pre-treatment variables to change over time.

$$Y_{ist} = \alpha + \underbrace{\sum_{k=-K}^{L}}_{k \neq \{E, -1\}} \gamma_k D_{st}^k + \sigma_i + \kappa_t + \eta BarBan_{ist} + \gamma X_i \times \kappa_t + \epsilon_{it} \tag{2}$$

Here, $D_{it}^k = 1(t - g_i = k)$, where g_i is the treatment (policy) year for unit i and $t - g_i$ is the relative time around the policy year. The never-treated states are included in the analysis and if unit i is never-treated then $D_{it} = 0$ for all t. The omitted categories include the year before the treatment and the earliest relative time, E. The two periods of the relative time are omitted to avoid problems of multicollinearity, which arises when all of the groups are eventually treated. Note that in the case with no never-treated units, the path of a fully dynamic event-study cannot be point identified as treatment timing g_i (subsumed by the unit FE) and calendar time t can perfectly generate the relative time R_{it}, where $R_{it} = t - g_i$ (See Borusyak, Jaravel, and Spiess (2021) for an excellent discussion regarding this issue). Only a handful of states remain untreated in suvery years 2004-2010. In other words, as the number of never-treated units are low, excluding the earliest period as well as the period prior to the treatment avoids concerns regarding multicollinearity.[18]

3.2 Issues with TWFE Estimator in context of cigarette taxation

The main issue with the TWFE estimator originates from the model assumption that the treatment effects are homogeneous across units and by the length of exposure to the treatment. While TWFE estimate is a weighted average of all possible 2 $\times$ 2 DD estimators in the sample, the weights may be negative (De Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021)). The negative weights can affect the TWFE estimate specifically when using the early treated units as comparison for units treated later in the sample.

To see this, we borrow explanation used in Roth et al. (2022). Using the Frisch-Lovell theorem express the TWFE estimator as:

[18] Although the number of never-treated group is quite large for the survey year 2015-2020, we still exclude the earliest time period for consistency as well as to comply with Sun and Abraham (2021) event-study approach.

$$\hat{\beta}^{TWFE} = \sum_i \sum_t \frac{(D_{it} - \hat{D}_{it})(Y_{it})}{(D_{it} - \hat{D}_{it})^2} \tag{3}$$

where $\hat{D}_{it}$ is the fitted value from the regression $D_{it} = \tilde{\sigma}_i + \tilde{\kappa}_t + \tilde{\epsilon}_{it}$, where $\tilde{\sigma}_i$ and $\tilde{\kappa}_t$ are unit and time fixed effects. Here, $Y_{it} = Y_{it}(0) + \tau_{it}(g)$; $Y_{it}(0)$ is the potential outcome for group g in absence of treatment and $\tau_{it}(g)$ is the treatment effect for group g at time t. The weights imposed on $\hat{\beta}^{TWFE}$ is directly proportional to $(D_{it} - \hat{D}_{it})$. However, the OLS estimation can predict $\hat{D}_{it}$ (treatment probability) greater than 1, in which case τ_{it} gets negative weights in equation 3 $(D_{it} - \hat{D}_{it} < 0)$. The negative weighting problem is more likely to affect the early treated units late in the sample. This can be explained by rewriting $\hat{D}_{it}$ as: $\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{D}$, where, $\bar{D}_i$ is the unit specific mean i, $\bar{D}_t$ is the mean specific to time t, and $\bar{D}$ is the grand mean across time and individuals.

For the early treated units $\bar{D}_i \approx 1$ and if all units are eventually treated by the end period of the sample, $\bar{D}_t \approx 1$ for t closer to the end of the sample. In this particular case, $\bar{D}_i + \bar{D}_t \approx 2$. Since, $\bar{D} < 1$ (i.e., there is a portion of untreated units or periods without treatment), $\hat{D}_{it} > 1$. This puts negative weights on the TWFE's decomposed 2×2 estimates. The negative weights are more likely to arise towards the end of the sample when the early treated units are compared to later treated units. While the negative weighting issue cancels out due to positive weights in other cases under homogeneous treatment effects, such is not the case when effects are heterogeneous.

The issue of negative weighting problem can also be seen through the lens of Goodman-Bacon (2021). Following his study, the TWFE estimate $\beta^{TWFE} = VWATT + VWCT - \Delta ATT$, where VWATT is the variance weighted average treatement effect for the treated, VWCT is the variance weighted common trends, and ΔATT refers to the change in treatment for the early treated group over the sample period (say, later vs. early period). In case of a dynamic treatment effect, $\Delta ATT \neq 0$, which exerts a negative weight on β^{TWFE}. In cases when the magnitude of ATT effects are increasing over time, the incidence of negative weighting problem is proportional to the change in the magnitude of the treatment effects.

How may this issue affect the TWFE estimate in case of treatment determined by increases in cigarette taxes? Depending on the sample period selected by the researcher, eventually almost all the units may receive treatment. For instance, if the choice of sample period is 1998-2010, all states will have eventually received treatement by the end year of the sample. Between 2004-2010, 38 states increased cigarette taxes. For the early treated units, $\bar{D}_i = 1$ and $\bar{D}_{t=2010} \approx 0.76$. Given that the timing of treatment varied across units, $\bar{D} < 0.76$. This means that $\hat{D}_{it}$ may be greater than 1 towards the end of the sample period, which exerts negative weight in equation 3. The negative weighting issue is less likely to be problematic when using the sample from 2015-2020 as only 18 states increased cigarette taxes during this period.

The presence of heterogeneity not only raises questions regarding the validity of the TWFE estimate from a static model as shown in equation 1 but also affects the canonical event-study estimates in equation 2. Sun and Abraham (2021) (SA) provide formal evidence to highlight problems regarding the canonical event-study as depicted in equation 2. The SA study highlights that even when assumptions of parallel trends and no anticipation effect are invoked, under heterogeneous treatment effects across cohorts (defined by the treatment timing), the event-study coefficient for a specific relative time l will be contaminated due to comparisons across other relative time periods ($l \neq l'$) and the excluded period. As such, the pre-treatment event-study indicators can capture treatment effects from the post-treatment periods, which would contaminate the pre-treatment estimates. This hinders a valid assessment of pre-treatment estimates to inform judgement regarding the parallel trends. Sun and Abraham (2021) propose an interaction-weighted estimator. First, the relative time indicators are interacted with group indicators (defined by treatment timing) using a TWFE specification to estimate dynamic average treatment effect per group.[19] Next, the group-time estimates are aggregated using weights defined as the sample share of each group at the given relative time.

3.3 Callaway and Sant'Anna Estimator - An alternative to TWFE

This study uses Callaway and Sant'Anna (CS) estimator following Callaway and Sant'Anna (2021), which allows for heterogeneous treatment effects across units and time. The CS estimator identifies

[19]To define units based on their treatment time, SA use the terminology "cohort", while CS use "group." SA's approach is similar to CS's approach when no additional covariates are included in the model i.e., when unconditional parallel trend assumption is invoked.

group-time treatment effect parameter for units that are first treated at time g (hence, group g) and estimated in calendar time t. Since the estimator uses never-treated (or not-yet-treated) units as the comparison group to identify the heterogeneous treatment effects, it provides a cleaner comparison. Thus, the CS estimator is not suseptible to issues associated with the TWFE model. The target estimand is:

$$ATT(g,\ t) = E(Y_t(g) - Y_t(0)|G_g = 1) \tag{4}$$

where, $Y_t(g)$ is the outcome of group g at time t, $Y_t(0)$ is the counterfactual, and G is a set of all possible treatment timing groups.[20] Under the following assumptions: i) unconditional parallel trend assumption which states that outcome of units treated at time g would have followed the same path as the untreated units in absence of the treatment, *i.e.*, $E(Y_t(0) - Y_{g-1}(0)|G_g = 1) = E(Y_t(0) - Y_{g-1}(0)|C = 1)$ where C is a set of comparison groups, and ii) no anticipation assumption, *i.e.*, $E(Y_t(g)|G_g = 1)$=$E(Y_t(0)|G_g = 1)$ for $t < g$, the ATT estimator is:

$$A\hat{T}T(g,\ t) = \frac{\sum_i (Y_{i,t}.1(G_i = g) - Y_{i,g-1}.1(G_i = g))}{\sum_i 1(G_i = g)} - \frac{\sum_i Y_{i,t}.1(G_i = C) - Y_{i,g-1}.1(G_i = C)}{\sum_i 1(G_i = C)} \tag{5}$$

where, $1(G_i = g)$ is a binary variable indicating whether a unit i is first treated in period g. Similarly, $1(G_i = C)$ is an indicator for untreated group. The assumption of no treatment anticipation allows using a period prior to the reform $(g-1)$ as the reference. Additionally, following the unconditional parallel trend assumption, the estimand assumes that the path of outcome for group g would have evolved similar to comparison group $(G_i = C)$ from period $g-1$ to t. The choice of comparison group as shown in equation 5 particularly refers to the never-treated units. Similarly, not-yet-treated units at time t can be validly used as the comparison group under the parallel trend and no anticipation assumptions. Note that equation 5 mimics the 2×2 DD framework where the

[20] Note that this is just an extention of the 2×2 DiD target estimand: $ATT = E(Y_2(D = 1) - Y_2(D = 0)|D = 1)$, where $D \in \{0,\ 1\}$ refers to the treatment status.

two groups are defined as g and C and the two time periods are $g-1$ vs. t.[21]

However, it is questionable whether the unconditional parallel trend assumption is likely to hold when evaluating smoking outcomes as states that are more likely to increase cigarette taxes may also have relatively high anti-smoking sentiments. It is likely that smoking outcomes would have evolved differently across states with different levels of smoking sentiments in absence of the treatment, violating the unconditional parallel trend assumption. In this case, conditional parallel trend assumption may be more plausible given that one observes pre-treatment smoking sentiments across states.

An important aspect of the CS method is that covariates can be flexibly incorporated in the analysis. This allows estimating conditional difference-in-differences which assumes that parallel trends hold between treated and untreated units with the same covariates. One can use the inverse probability weighting (IPW) (Abadie (2005)), outcome regression (OR) (Heckman, Ichimura, and Todd (1997)), or doubly robust (DR) (Sant'Anna and Zhao (2020)) methods to recover the DiD parameters, while flexibly accounting for the pre-treatment covariates. Although all three approaches recover the same parameter under the conditional parallel trend assumption from an identification standpoint, this study relies on DR approach due to its additional robustness property. As discussed in detail in Sant'Anna and Zhao (2020), DR estimator is preferred from an empirical perspective (compared to the other two methods) as it enjoys an additional robustness property that only requires either propensity score or outcome regression specification to be correct for valid estimates. The pre-treatment locality specific covariates that are accounted for include a measure of anti-smoking sentiment in 1998, unemployment rate (in 2000 and 2010 for the 2004-2010 and 2015-2020 samples, respectively), the change in proportion of current smokers between 1998-1999 and 2001-2002, and the log of area specific population.

The estimation of DR approach can be grouped into two steps. First it estimates the probability that a unit falls in group g, denoted as $\hat{p}_g(x)$ as well as models the outcome evolution using the approach suggested in Heckman, Ichimura, and Todd (1997). The second step uses $\hat{p}_g(x)$ as the weights and also adjusts the predicted changes in the outcome for the treated group in

[21] Units treated in the first year of the sample period have no pre-treatment comparison. Hence, these units will not contribute to the estimation.

absence of treatment. The latter is obtained by first estimating the conditional expectation for the untreated units using their covariates X_i. The estimates from this regression are used to predict outcomes for the treated group by using X_i specific to the treatment group given as $\hat{m}_{g,t}(X_i) = \hat{E}[Y_{i,t} - Y_{i,g-1}|D_i = 0,\ X_i]$. As such this method incorporates both IPW and OR methods. The DR estimator is given as:

$$A\hat{T}T(g,\ t) = \frac{1}{N}\sum_i \left[\left(\frac{1.(G_i = 1)}{\frac{\sum_i 1.(G_i=g)}{N}} - \frac{\frac{\hat{p}_g(X)1.(G_i=C)}{1-\hat{p}_g(X).1(G_i=C)}}{\frac{1}{N}\sum_i \frac{\hat{p}_g(X)1.(G_i=C)}{1-\hat{p}_g(X).1(G_i=C)}}\right)(Y_{i,t} - Y_{i,g-1} - \hat{m}_{g,t}(X))\right] \tag{6}$$

Here, the units in comparison group are weighted inversely according to their probability of being untreated. This is given by weights being proportional to $\frac{\hat{p}_g(X)}{1-\hat{p}_g(X)}$ for untreated units. Intuitively, units in untreated group that have low probability of getting treated can be very different from the units in treatment group. Hence, these units receive lower weights.

Following the recovery of disaggregated parameters, $A\hat{T}T(g,\ t)$, these can be aggrregated to summarize the treatment heterogeneity by using the appropriate weights. One aggregation of interest for this study is to evaluate how the treatment effect varies with the length of exposure to the treatment. This portrays dynamic treatment effects similar to the canonical event-study approach from equation 2 but by relaxing the assumption that effects are homogeneous across units. To evaluate how the treatment effect varies with the length of exposure to the treatment, an aggregation approach given below is adapted.

$$\theta_{es}(e) = \sum_{g\in G} w_e^{es}(g,t)ATT(g,\ g+e) \tag{7}$$

where the weight, $w_e^{es}(g,t) = P(G = g|G + e \leq T)$. Here, T is given as the number of periods in the sample and e is the relative time period (the number of periods before or following the treatment period, i.e., $e = t - g$). In other words, weights are just the proportion of units treated in time g when measured at the relative time e. $\theta_{es}(e)$ documents the effect of the treatment by length of exposure to the treatment, which is contextually similar to the event-study parameters from the TWFE in equation 2. The standard errors are obtained using a multiplier bootstrap approach that is robust to multiple-testing problems unlike the pointwise standard error.

4 Diagnostic checks and Simulation

Before discussing the main findings, the study inspects three separate forms of diagnostic tests to assess the performance of the TWFE model.

Diagnostic check I. A straightforward and informative diagnostic check to assess the performance of the TWFE estimate is acquired from equation 3. Note that the weights given during the estimation of $\hat{\beta}^{TWFE}$ are directly proportional to $(D_{it} - \hat{D}_{it})$.[22] An illustration of $(D_{it} - \hat{D}_{it})$ for each group defined by the treatment timing across the calendar time t will provide a simple check of how the weights are being allocated in estimation of the TWFE estimate.

Diagnostic check II. How does the TWFE estimates compare to the true ATT (treatment effect or treatment effects) among the samples used in this study? Using the actual variation in cigarette taxes throughout the sample captured by D_{it} in equation 1, the study uses simulation exercises to gauze the performance of the TWFE model. Here, a true treatment effect is imposed and compared with the TWFE estimate under the following conditions: *i*) homogeneous treatment effect, *ii*) heterogeneous treatment effects across units, and *iii*) heterogeneous treatment effects by time. Using the actual variation in cigarette taxes for simulation allows an explicit assessment of the TWFE estimate in cases of heterogeneity. The details regarding the setup of the simulation exercise are discussed in Section 3.3.

Diagnostic check III. Next, to gauge the role of *bad comparison* units in TWFE setting when using early treated group in comparison to late treated group, the analysis uses the decomposition of TWFE estimate provided in Goodman-Bacon (2021) and report aggregate weights given to the following categories: *i*) early treated group (as the treated group) vs. later treated group (comparison), *ii*) later treated vs. always treated, *iii*) later vs. early treated, and *iv*) treated vs. untreated. Categories *i*) and *iv*) provide "clean" comparison, while the other two may be

[22] For each group defined by the treatment timing, $(D_{it} - \hat{D}_{it})$ varies over time but is invariant across units at a given calendar time t within the group.

contaminated in cases of treatment heterogeneity. All the diagnostic checks are conducted separately for the 2004-2010 and 2015-2010 sample periods.

Results from diagnostic checks. Figure 2 plots the character of weights (positive vs. negative) categorized by the assessment of $D_{it} - \hat{D}_{it}$ for the 2004-2010 and 2015-2020 samples in panels A and B, respectively **(diagnostic check I)**. Panel A shows that units treated in 2005 (*group* $g = 2005$) exert a negative weight starting from the year 2008. Similarly, units in group 2006 impose negative weights in 2010. Given that the effects of cigarette tax incidence increases in magnitude over time, Panel A in Figure 2 suggests that the TWFE estimate will be suppressed towards 0.

Next, Figure 3 presents results from simulation exercises pertaining to **(diagnostic check II)**. Panel A indicates that the TWFE model performs well during the case of homogeneous treatment effect. The mean of the replicated estimates from the simulation is close to the true effect. In contrast, the TWFE model does a poor job when the effects are heterogeneous in cases *ii*) and *iii*). The magnitude of the mean of the replicated TWFE estimates are three-fourth and two-fifth the size of the true effects in case *ii*) and case *iii*). The simulation exercise points out that the performance of the TWFE model deteriorates in case of treatment heterogeneity for the 2004-2010 sample.

Table 3 presents results from the Goodman-Bacon (2021) decomposition to understand how the early treated group might affect the TWFE estimates by forming *bad comparison* **(diagnostic check III)**. The table summarizes the weights given for all possible 2×2 difference-in-differences estimates into four categories. As previously mentioned, since the effects of treatment may vary with the length of exposure, the problem with TWFE arises when comparison of the early treated group forms "bad comparison" for units treated later on in the sample period. This group (category *iii*) contributes 32% of weight to form the TWFE estimate for the 2004-2010 sample. While it carries the highest of weight, the magnitude of estimate pertaining to this group is the lowest among the four groups. This is explained by a large number of states (38 states) eventually being treated in the earlier sample, with a significant portion of states receiving treatment in the later years of the sample period. In comparison, the weight given for later vs. earlier treated group (category *iii*) is only 4.7% for the 2015-2020 sample.

Next, the difference in weights for each group defined by the treatment timing when serving as treated vs. control units are shown in Figure 4 using the triangular markers along with the sample share for each group in solid circles. I replicate the exercise after dropping always- and never-treated units to focus only on the variation due to the treatment timing. Units receiving treatment in 2005 and 2010 contribute disproportionately more as control units rather than treated units compared to units treated towards the middle of the sample period. In fact, units treated in 2005 serve more as control units than treated while forming the TWFE estimate. This further highlights the problem of early treated group acting as comparison for units treated later in the sample ("bad comparison group"), which suppresses the TWFE estimate towards zero in case of treatment heterogeneity by time (i.e., the treatment effect increases by time). In the 2015-2020 sample, units treated in 2018 serve disproportionaly more as treated than control units compared to any other groups. In the case of heterogenenous treatment effects across units, if the 2018 group have treatment effect that is lower in magnitude compared to other groups, then the TWFE estimate will be suppressed towards zero.

Simulation exercises to assess the performance of CS estimator. Subsequently, using simulation exercises are designed to evaluate the performance of the CS estimator based on the treatment variation defined in Table 1, considering treatment effects that are: *i*) homogeneous, *ii*) heterogeneous across units, and *iii*) heterogeneous by relative time.[23] The results from simulations are presented in Figures 12, 13 and 14 in the Appendix. The mean of the replicated estimates are similar to the true effects across all three cases. These simulations validate that the CS estimator outperforms the TWFE model in scenarios involving treatment heterogeneity in the context of cigarette taxation.

In summary, the diagnostic checks reveal that using already treated group as the comparison group is problematic in presence of treatment dynamics, particularly for the early sample period. Furthermore, the CS estimator demonstrates the ability to effectively capture treatment dynamics based on the variation in treatment as well as the sample used in the study.

[23]As previously mentioned, the setup governing the simulation exercise is discussed in the Appendix Section 3.3.

5 Results

5.1 TWFE Results

The results from TWFE models are presented in Table 4 when using the percent of current smokers as the dependent variable. Columns 1-3 and 4-6 refer to the 2004-2010 and 2015-2020 samples; Panel B drops units that received treatment in the very first period of the sample (always treated) since they do not contribute to the identification in TWFE models as treatment units.[24] Columns 1 and 4 provide results from the parsimonious specification that only includes a dummy for cigarette tax change along with the year and MMSA (metropolitan micropolitan statistical area) fixed effects. Columns 2 and 5 include the percent of population living under smoking ban in bars, while Columns 3 and 6 additionally include a vector of pre-treatment characteristics including the locality specific log of population, unemployment rate in 2000 (for the 2000-2010 sample) or 2010 (2015-2020 sample), a measure of anti-smoking sentiment in 1998, and the proportion change in current smokers between 1998 and 2001. These time invariant covariates are interacted with the year dummies as shown in equation 1.

Table 4 shows that the coefficients on the treatment variable are negative across all specifications and both panels, suggesting that cigarette tax increases are negatively associated with the prevalence of smoking. Column 1 suggests that an instance of cigarette tax increase on average is associated with a reduction in prevalence of current smoker by 0.6 percentage points and the coefficient is statistically significant at the 5 percent level. Next, including the percent of population living under smoking ban in bar (Column 2) and additional covariates (Column 3) do not affect the magnitude on the treatment indicator. The coefficients on the treatment indicator presented in Panel B are similar to Panel A. The proportion of current smokers decreased by 4.25 percentage points between 2004 and 2010.[25] If we were to allow for causal interpretation on the treatment depending on the TWFE estimate with the largest magnitude, it would suggest that on average 14 percent of the reduction in current smoker can be attributed to instances of tax increment

[24]The identification in TWFE model relies on within-unit variation in treatment over time. However, always treated group poses a challenge since units in this group lack variation in treatment. They are rather falsely utilized as comparison units in the model, even though their untreated potential outcomes are never observed.

[25]Author's calculation using the BRFSS SMART data.

between 2004-2010. It is evident that tax increases in the relatively recent sample years (2015 to 2020) are also negatively associated with the proportion of current smokers as shown in Columns 4-6. However, the coefficients on the treatment indicator are imprecisely estimated as well as the magnitude of the estimates are relatively smaller to the early period (2004-2010) across all specifications.

Additionally, we run the TWFE specification with a control for whether cigarette taxation was implemented after July during the implementation year. The findings shown in the Appendix section, Table 6, are similar to the main findings. Next, we drop the implementation year and use it as an adjustment period. The TWFE estimates from this approach, shown in the Appendix Tables 7, are slightly larger than the estimates reported in Table 4. This can be explained by the gradual increase in the impacts of cigarette taxation over time, as evidenced by the event-study.

There are two main issues with the TWFE estimates presented in Table 4. First, issues with the TWFE comes from the *negative weighting* problem as highlighted in section 4 using diagnostic checks. Second, although specifications in Columns 2, 3, 5 and 6 invoke conditional parallel trend assumption, the time invariant pre-treatment covariates enter the model specification through an interaction with the year fixed effects. Given that the treatment effects vary within different values of covariates, introducing covariates linearly may not be appropriate (see Meyer (1995)).

5.2 Results allowing for heterogeneity in treatment effects by length

To allow for dynamic treatment effects by the length of exposure to treatment, I first estimate the canonical event-study specification given by equation 2. In addition, I also report estimates proposed in Sun and Abraham (2021) to correct for the possible contamination that might affect the canonical event-study estimates. The results from the canonical event-study as well as SA design are shown in Figures 5 and 6 for the sample periods 2004-2010 and 2015-2020.[26] Panels A and B refer to results from the parsimonious specification and with additional covariates as controls, respectively.

The findings from both the canonical event-study and SA designs presented in Panel A, Figure 5,

[26]The canonical event-study estimates are reported in Tables 8 and 9.

show reductions in the prevalence of being a current smoker immediately starting from the period of cigarette tax implementation. The effects are double the size in magnitude a year following the tax implementation after which the magnitude of the estimates increase gradually in size. The average of the event-study estimates in the post period shown by the horizontal black dotted line using the canonical design is slightly below -1, whereas the red dotted line refers to the TWFE estimate shown in Table 4.[27] The TWFE estimate is around 46% as large as the average of event-study estimates in the post period. While estimates from the event-study design (that allows for heterogeneity over time) suggest that 23% of the reduction in prevalence of smoking between 2004-2010 is attributed to tax increases, the TWFE estimate accounts for only 14% of the reduction. Moreover, the pre-treatment estimates from both canonical as well as SA design are statistically indistinguishable from zero (p-value = 0.369 from the Wald test for joint nullity in the case of the canonical design), which supports the parallel trend assumption governing the difference-in-differences framework. The results presented after controlling for covariates in Panel B, Figure 5, exhibit similar patterns as observed in Panel A. The diagnostic checks and Goodman Bacon decomposition, as discussed in section 4, offer compelling insights as to why TWFE estimates exhibit a bias towards 0 in the context of cigarette taxation in United States.

The dynamic results referring to the later sample (2015-2020), shown in Figure 6, also demonstrate heterogeneity over time. The coefficients gradually increase in magnitude following the year of tax implementation and the size of the coefficients are below -1 in the third and fourth relative years. This pattern is similar across both panels representing specifications with and without additional covariates. The TWFE estimates shown by the red dotted line is only slightly less in magnitude compared to that of the average of post treatment event-study estimates from the canonical design.

5.3 Results from the CS Estimator

Figures 7 and 8 show the group-time treatment estimates aggregated to show the average treatment effects on the treated ($\hat{ATT}$s) by the length of exposure to the treatment using the aggregation

[27] The event-study estimates from the canonical design are presented in tables 8 and 9.

scheme shown in equation 7. Figures 7 and 8 refer to years 2004-2010 and 2015-2020, respectively. Panels A and B use never-treated units as the comparison group, whereas Panels C and D use not-yet-treated units. Panels A and C show the results without additional covariates, while Panels B and D flexibly control for the pre-treatment covariates (the log of local area population, unemployment rate, a measure of anti-smoking sentiment in 1998, and the percent change in current smokers between 1998-2001) using the Doubly Robust approach. All sub-figures report the 95% simultaneous confidence bands obtained from the bootstrapped standard errors.

The estimates in Figure 7, Panel A, show that the percentage of current smokers decrease following the treatment implementation. The figure does not provide any evidence of differential pre-treatment trend. The pre-treatment estimates fluctuate around zero and are statistically insignificant. The effect during the year of tax increase is negative but statistically insignificant at the conventional levels. But the effects are pronounced starting from the second year of the treatment, where the magnitude of $A\hat{T}T$ is below 1 and the effect is statistically significant at the 5 percent level. There is a gradual decline in the effect size as the length of treatment exposure increases. The $A\hat{T}T$ is close to -2 following 5 year of exposure to the tax incidence. These $A\hat{T}T$ effects are quite similar to the pattern as well as the size of the event-study estimates shown in Figure 5. Moreover, the dynamic effects of the treatment estimated using the CS estimator remain consistent across all panels, including: *i*) additional covariates when using never-treated as comparison (Panel B), *ii*) not-yet-treated group as comparison instead of only never-treated units (Panel C), and *iii*) accounting for covariates when using not-yet-treated group as comparison (Panel D).

Panel A, Figure 8, shows the impacts of tax incidence in the later period (2015-2020). Consistent with the identification assumption, I find no evidence of systematic differences in trends prior to the treatment. Following the treatment, the size of $A\hat{T}T$ gradually increases in magnitude and the effect is around -1.5 percentage points after 5 years of exposure to the treatment. Although a clear break in trend is visible in the post-treatment era, none of the $A\hat{T}T$s are statistically significant at the 5% level. Hence, these findings should be interpreted with caution.

The overall summary of the group-time average treatment effects using the CS estimator are provided in Figure 9 as point estimates. Panels A and B refer to the years 2004-2010 and 2015-2020, respectively. The 95% confidence bands are plotted around the average treatment effects. The first $\hat{ATT}$ *(01. nt(no controls))* is generated from using never-treated units as the comparison group and without additional covariates. The second $\hat{ATT}$ is generated from the aggregation of group-time effects that are obtained by using DR approach with covariates. The third *(03. nyt(no controls))* and fourth *(04. nyt(controls))* ATT estimates are obtained similarly to the first and second estimates but after using not-yet-treated units as comparison instead of only never-treated units. The TWFE estimate with the largest magnitude from Table 4 are show by the horizontal line for reference across the two panels.

Panel A shows that the overall $\hat{ATT}$ across four different estimation approaches are similar in magnitude. More importantly, the magnitude of overall $\hat{ATT}$ is more than two times the size of the TWFE estimate, shown by the red dotted horizontal line. This clearly demonstrates the difference in the treatment effect intensity generated from the TWFE model vs. group-time treatment approach when using the 2004-2010 sample.

The magnitude of overall $\hat{ATT}$s in the 2015-2020 sample, as shown in Panel B, are similar across different approaches but are only about 65 percent of the size of $\hat{ATT}$ demonstrated in Panel A. In other words, the findings indicate that the efficacy of cigarette taxation as a tool to improve smoking outcomes has lowered in recent periods. The magnitude of the overall point estimates from the group-time treatment effects are still slightly larger than the TWFE estimate. It has to be noted that none of the overall $\hat{ATT}$s in Panel B are statistically significant at the 5 percent level, although by small margins.

What could explain the decline in the effectiveness of cigarette taxation in recent years? Firstly, the overlap of tax hikes in both the early (2004-2010) and later (2015-2020) sample periods across eleven out of eighteen states may have altered the composition of current smokers between these periods. This could result in an increased proportion of price-insensitive smokers in states treated later in the sample period, as price-sensitive smokers may have already reduced or quit smoking in response to early tax increases. Additionally, considering the treatment dynamics defined by the length of treatment, whereby MSAs within twenty-seven states served as treatment units in the early sample period now act as control units in the later sample, may bias the treatment effects towards zero in the later sample.

Robustness checks. One issue when estimating the dynamic treatment effects by the length of exposure to the treatment is that the composition of units across relative time bins changes. For example, ATT for units treated in 2009 and 2010 will not be identified for the relative time periods greater than 1 and 0 in the 2004-2010 sample period, respectively. To address the issue of changes in composition of groups, we re-estimate CS approach by including only the units balanced in the relative time period. Specifically, we include units that are at least treated for 3 and 2 additional periods following the treatment year for 2004-2010 and 2015-2020 samples. The results from this exercise are shown in Figures 15 and 16 in the Appendix. The dynamic estimates as well as overall $A\hat{T}T$ for the sample period 2004-2010, shown in Figure 15, are similar to the main results (Figures 7 and 9), suggesting that the results are not driven by compositional changes of units in relative time. For the 2015-2020 sample, while the estimates increase in magnitude following the treatment year, they are statistically indistinguishable from zero.

6 Discussion on elasticity estimates (extensive margin)

Given the findings that the effects of tax incidence are heterogenous on relative time, the elasticity equation should be adjusted to account for: i) treatment heterogeneity; and ii) different realizations of relative time across groups (imbalance on relative time). To highlight the latter point, units treated in year 2005 in the 2004-2010 sample will have 6 periods of relative time, while units treated in 2006 will have only 5. To consider the two aforementioned concerns, I focus on group specific elasticity estimate (conditional on group $G_i = g$), E_g. This is written as:

$$E_g = \sum_{k=0}^{K} \theta_k^g \frac{A\hat{T}T_k^g}{Y_{g-1}} \times \frac{tax_{g-1}}{\Delta tax} \tag{8}$$

$$\theta_k^g = \frac{1}{\sum_t^T \sum_{k=0}^{K} 1(t-g=k)} \tag{9}$$

$t \in \{t_0, t_1, ...T\}$ and $g \subseteq t$. k is the relative time and the base year is taken as the year prior to the treatment year $(g-1)$. $\hat{ATT}^g_k$ is the group-relative-time specific treatment effects estimated using the CS estimator. θ^k_g is the weight given to the treatment effect pertaining to the relative time k.

The elasticity estimates for the extensive margin are shown in Figure 17. Panel A shows that 1 percent increase in cigarette tax in years 2004 and 2005 reduced the percent of current smokers by 0.07 and 0.16 percent, respectively. The estimates for the groups treated in 2007-2010 are relatively smaller. This may be explained by the fact that units treated later on in the sample have a fewer periods of relative time, which disables the treatment dynamic to be fully realized given the data structure. Alternatively, the TWFE estimate with the largest magnitude yields an elasticity estimate of -0.025.[28]

Panel B shows that a percent increase in cigarette taxes for the group treated in 2016 reduced current smokers by 0.16 percent. This estimate is similar to that of the group treated in 2006 (in Panel A), with both groups having 6 years of relative time in the sample. In summary, the findings suggest that the elasticity estimates in the case of heterogeneity depends on the sample length and using the TWFE estimate to calculate elasticity can suppress the elasticity estimates to zero.

7 Conclusion

This study re-evaluates the role of cigarette taxes in curtailing smoking by using data from the Behavioral Risk Factor Survelliance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART) for the two sample periods: *i*) 2004-2010, and *ii*) 2015-2020. Although the topic of whether cigarette taxes can be used as an effective policy instrument to reduce smoking has been widely addressed, many of the studies conducted in the past decades rely on the two-way-fixed (TWFE) approach where the treatment is rolled out in a multiple-group and multiple-timing setting. While the TWFE approach reveals the average treatment effect on the treated in absence of heterogeneity across the treated units and the length of exposure to the treatment given that the parallel trend assumption is satisfied, such is not the case when the treatment effects are heterogenenous. This study incorporates recent advances that have been

[28]This is calculated as $\frac{\hat{ATT}_{TWFE}}{Y_{2004}} \times \frac{tax_{2004}}{avg.\ \Delta tax}$.

made in the multiple-group and multiple-timing difference-in-differences framework to comment on the average treatment effect of cigarette tax incidence by comparing the newer estimates with the TWFE estimate.

The results show that the largest TWFE estimate pertaining to 2004-2010 sample is less than half the size in magnitude to the overall average treatment effect on the treated obtained by using the group-time treatment effect ($A\hat{T}T$) approach following the Callaway and Sant'Anna (2021) estimator. While the TWFE point estimate suggests that 14% of the reduction in the prevalence of being a current smoker between 2004 and 2010 is explained by tax incidence, the overall group-time estimate shows that tax incidence contributes to over 25% of the reduction. Since many units were eventually treated by the end of the sample period 2004-2010, the diagnostic checks reveal the problem of negative weights associated with the TWFE approach. Moreover, the Goodman-Bacon (2021) decomposition reveals that 32% of the weight in forming the TWFE point estimate comes from the comparison between later treated units (as treated) vs. those treated earlier in the sample (as control). This is precisely the comparison one would like to avoid. The canonical event-study design and the design proposed by Sun and Abraham (2021) yield dynamic estimates similar to those obtained by following Callaway and Sant'Anna (2021), indicating that the magnitude of average treatment effect on the treated increases with the length of exposure to the treatment.

Only a handful of states were treated by the end of the later sample period (2015-2020). Hence, the issue of *bad comparison* group is not as critical in the later period as it is in the earlier sample. However, the magnitude of the overall treatment effect from the group-time setting is still slightly larger than that of the TWFE estimate. The comparison of $A\hat{T}Ts$ across two sample periods reveal that the the overall point estimate of $A\hat{T}T$ in the later sample period is only 65% of the size of $A\hat{T}T$ pertaining to the earlier sample period. The reduced effectiveness of tax incidence in more recent years may explained by increased price insensitivity due to the change in composition of smokers in later years.

These findings offer critical insights for policymakers and practitioners in the realm of cigarette taxation. Effective policy-making hinges on comprehensive cost-benefit analyses to gauge societal welfare. However, relying solely on estimates derived from traditional TWFE models may obscure the true effectiveness of cigarette taxation in the United States, leading to underestimated elasticity estimates. More accurate estimates are essential for informed decision-making and efficient resource allocation. For instance, in response to the observed decline in the effectiveness of cigarette taxation in recent years, policymakers may consider implementing additional measures, such as targeted awareness/control programs, to complement existing efforts and achieve desired public health outcomes (Farrelly, Pechacek, and Chaloupka (2003), Farrelly et al. (2008)). Furthermore, researchers and practitioners have demonstrated a multifaceted interest in exploring the broader impacts of cigarette taxes on various outcomes beyond smoking, including birth outcomes, alcohol and marijuana consumption, body mass index (BMI), and labor market outcomes. The outcomes elucidated in this study provide more precise 'first-stage' estimates, crucial for understanding the intricate role of cigarette taxation, as its effects on other outcomes often stem from its influence on smoking behavior.

Several limitations are worth mentioning. First, the study uses a binary form of treatment rather than different treatment intensities determined by the dosage of tax increases. While uncovering $A\hat{T}T$ based on the treatment defined by dosage relies on stronger assumptions than the case when the treatment is binary (Callaway, Goodman-Bacon, and Sant'Anna (2021)), the use of binary treatment does not allow us to evaluate the heterogeneous effects of the magnitude of tax increase. It is worth mentioning that this issue still prevails when using TWFE with continuous tax values. Second, using the BFRSS SMART data excludes some states from the analysis and only includes individuals residing in metropolitan or micropolitan areas. This may create problems of generalizing the findings to the entire population in America. Future studies can benefit from addressing these issues.

8 Figures and Tables

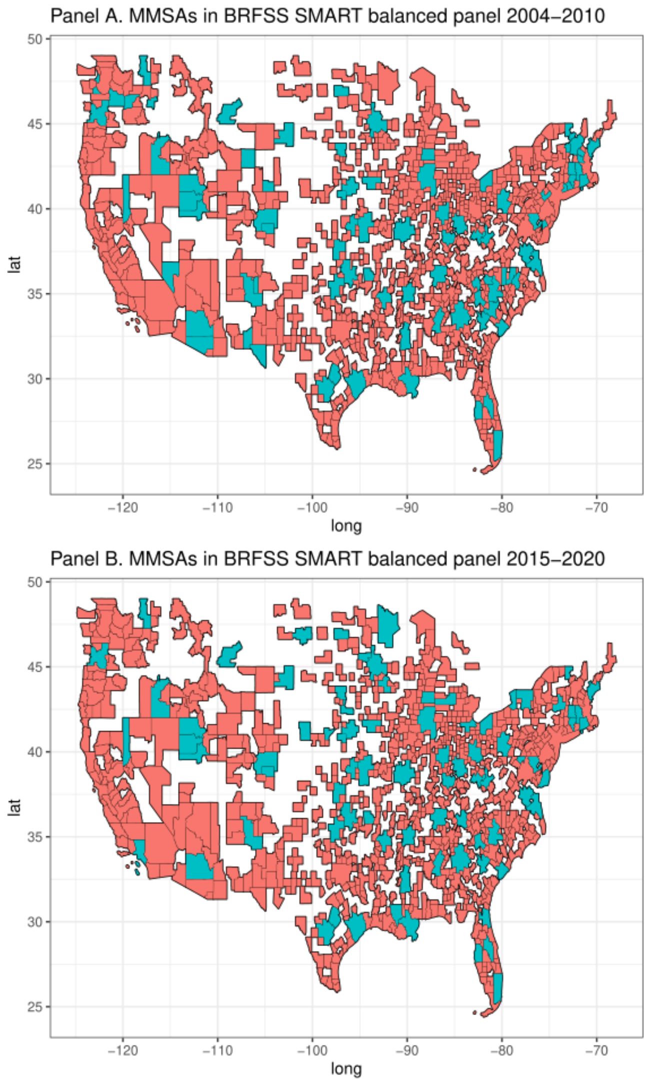

Figure 1: MMSAs in BRFSS SMART
Note: The figures show the map of Metropolitan/Micropolitan core-based statistical areas (CBSA), where the green polygons are MMSAs represented in the balanced panel of BRFSS SMART. There are 108 and 95 MMSAs for the sample years 2005-2010 and 2015-2020.

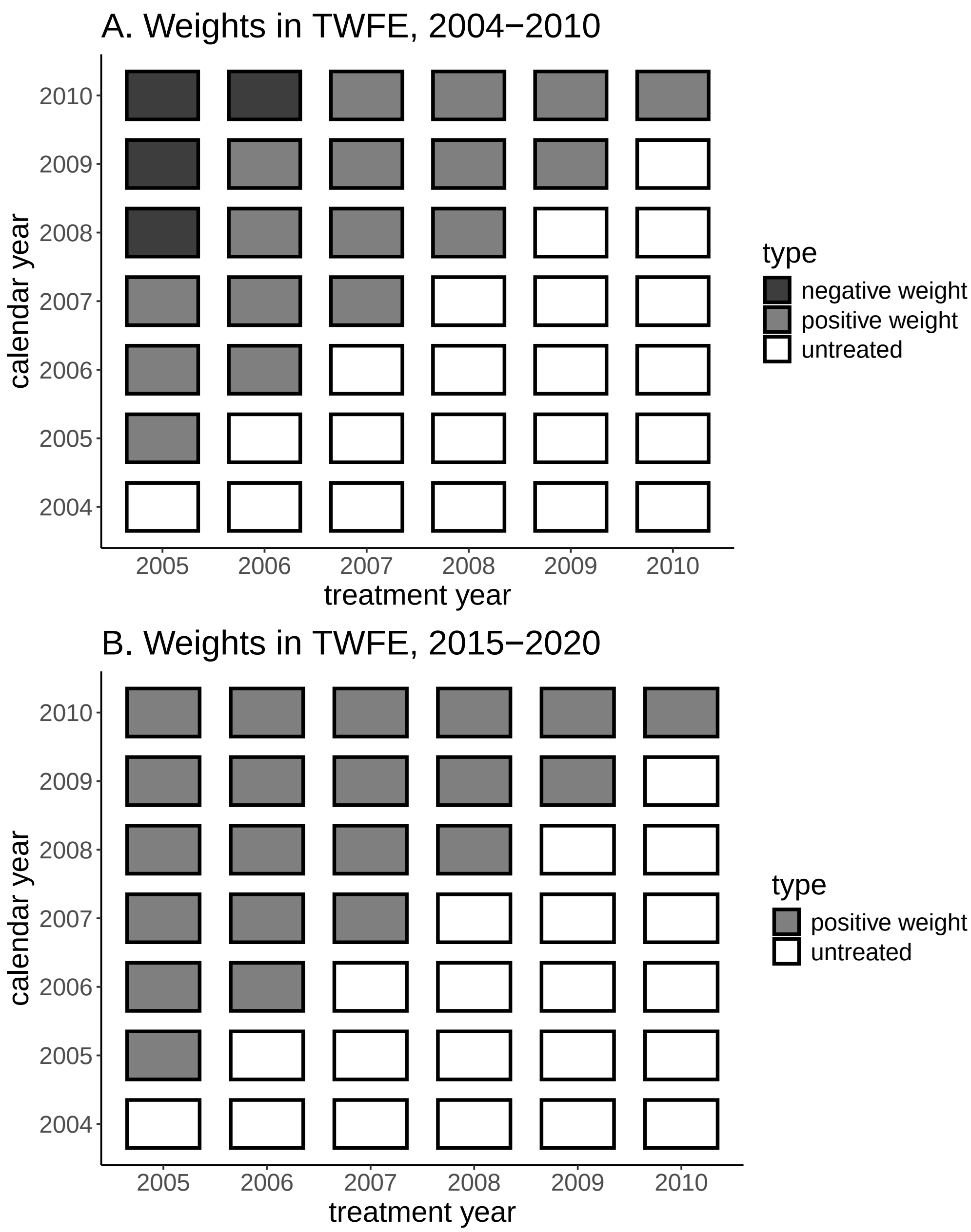

Figure 2: Weights in TWFE
Note: The figure shows the character of weights (negative, postive) that are used during the calculation of the TWFE estimate by the treatment year group throughout the years following the treatment in the sample period. Panels A and B refer to the 2004-2010 and 2015-2020 samples respectively.

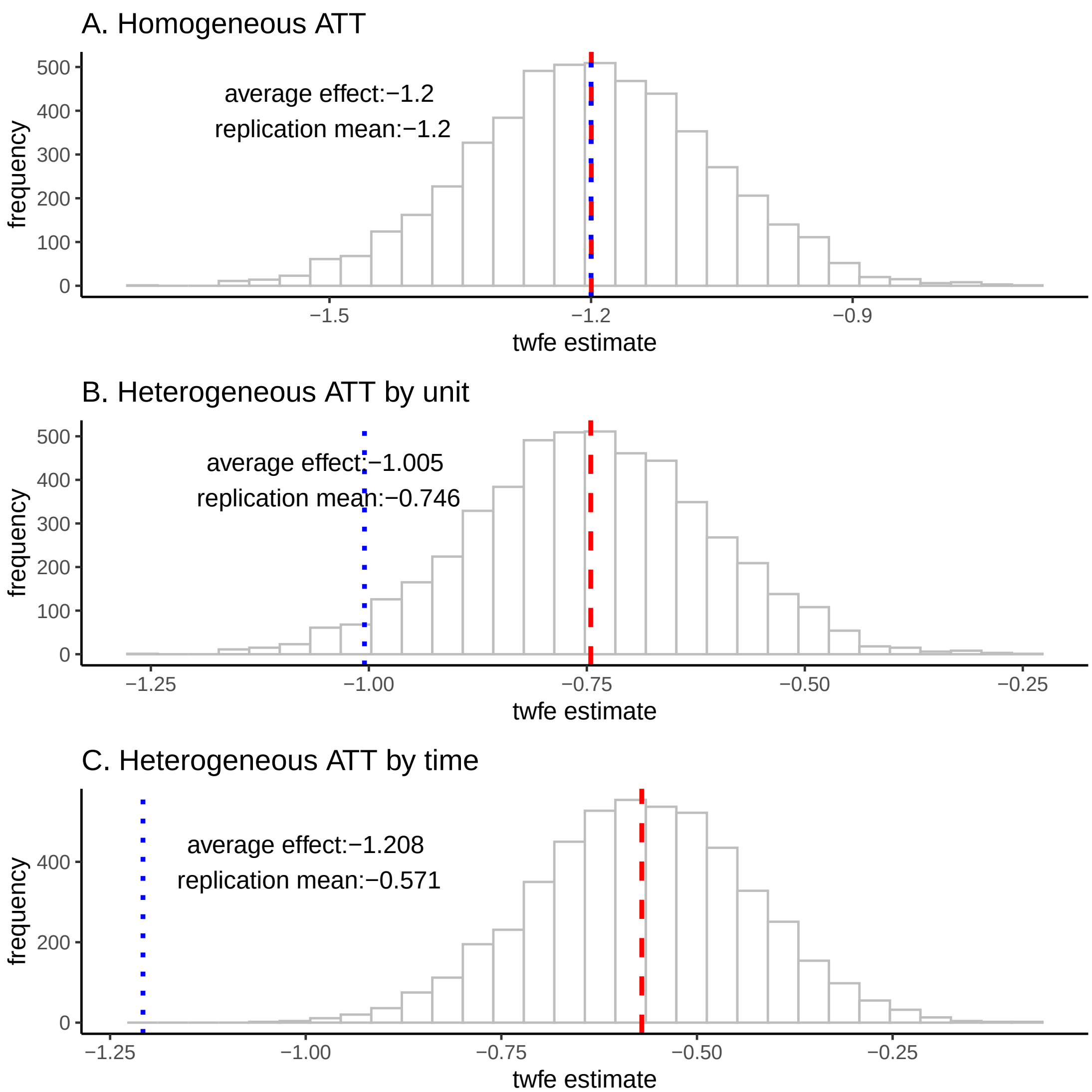

Figure 3: Simulation – Performance of TWFE
Note: The figure presents results from the simulation exercise that uses the variation in tax changes for the sample period 2004-2010. Figures 1, 2, and 3 present three different cases under the assumptions of: *i*) homogeneous treatment effect, *ii*) heterogeneous treatment effect by units, and *iii*) heterogeneous treatment effects by relative time. The simulation set up is discussed in the Appendix section 03.

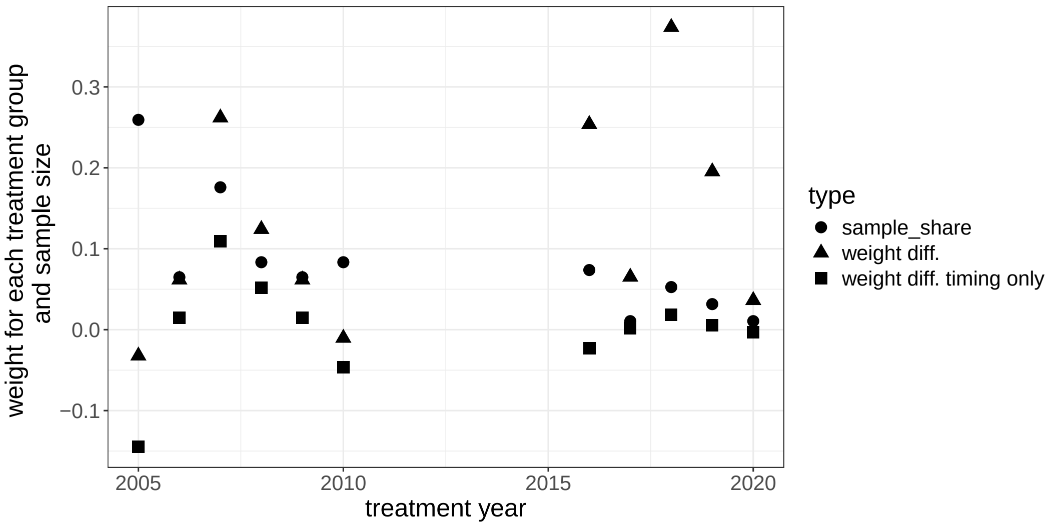

Figure 4: Treated/Untreated Weight Difference
The figure shows the difference in weight when each group based on the treatment timing acts as the treated vs. comparison group during the TWFE estimation following the decomposition shown in Goodman-Bacon (2021). The square markers represent results after excluding always treated as well as never treated groups and focus solely on variation due to the treatment timing. The sample size for each group are shown in solid circles.

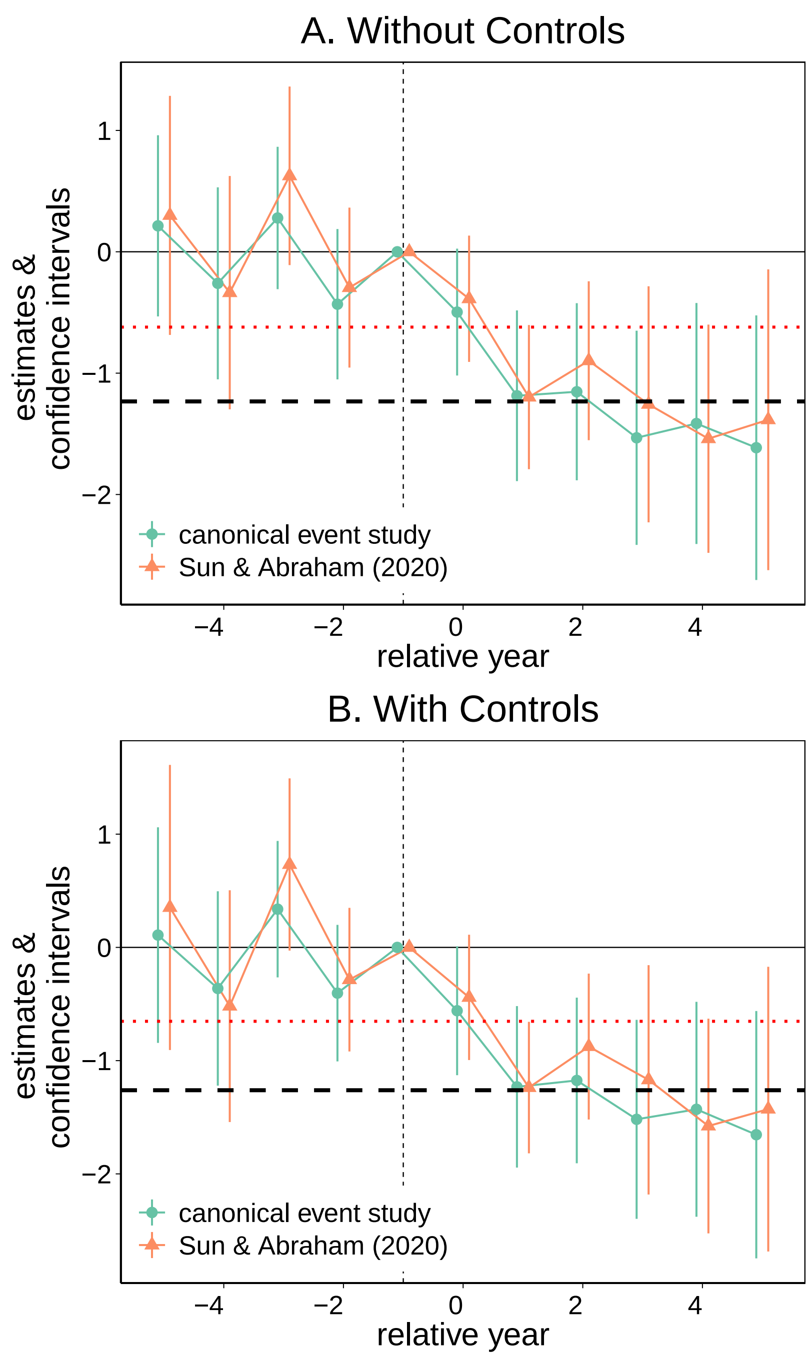

Figure 5: Event Study Estimates (2004-2010 sample)

Note: The figures show findings from the event-study analysis based on the TWFE as well as SA (2020). Panel A uses the parsimonious specification, while Panel B includes controls for the percent of state population living under smoking ban in bars along with the pre-treatment variables of the locality (*i*. log of population measured in 2000, *ii*. the change in current smokers between 1998 and 2001, *iii*. a measure for anti-smoking sentiments in 1998, *iv*. unemployment rate) interacted with year dummies. The vertical bars represent the 95 percent confidence intervals constructed from standard errors clustered at the state level. The red dotted lines represent the largest TWFE estimate from Table 4, while the black line is the average of the post-treatment event-study estimates obtained from the canonical model.

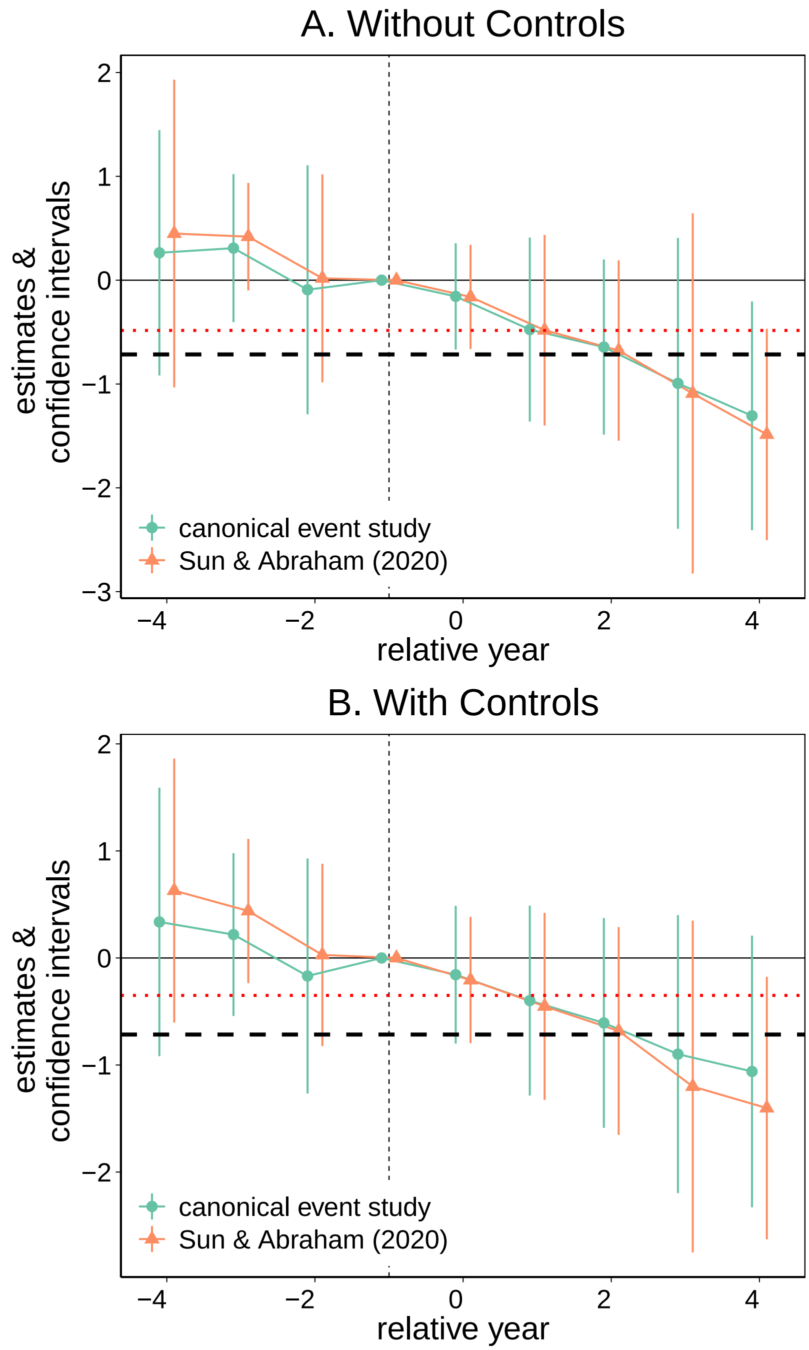

Figure 6: Event Study Estimates (2015-2020 sample)
Note: The figures show findings from the event-study analysis based on the TWFE and SA (2020). Panel A uses the parsimonious specification, while Panel B includes controls for the percent of state population living under smoking ban in bars along with the pre-treatment variables of the locality (*i*. log of population measured in 2010, *ii*. the change in current smokers between 1998 and 2001, *iii*. a measure for anti-smoking sentiments in 1998, *iv*. unemployment rate) interacted with year dummies. The vertical bars represent the 95 percent confidence intervals constructed from standard errors clustered at the state level. The red dotted line represents the largest TWFE estimate from Table 4, while the black line is the average of the post-treatment event-study estimates obtained from the canonical model.

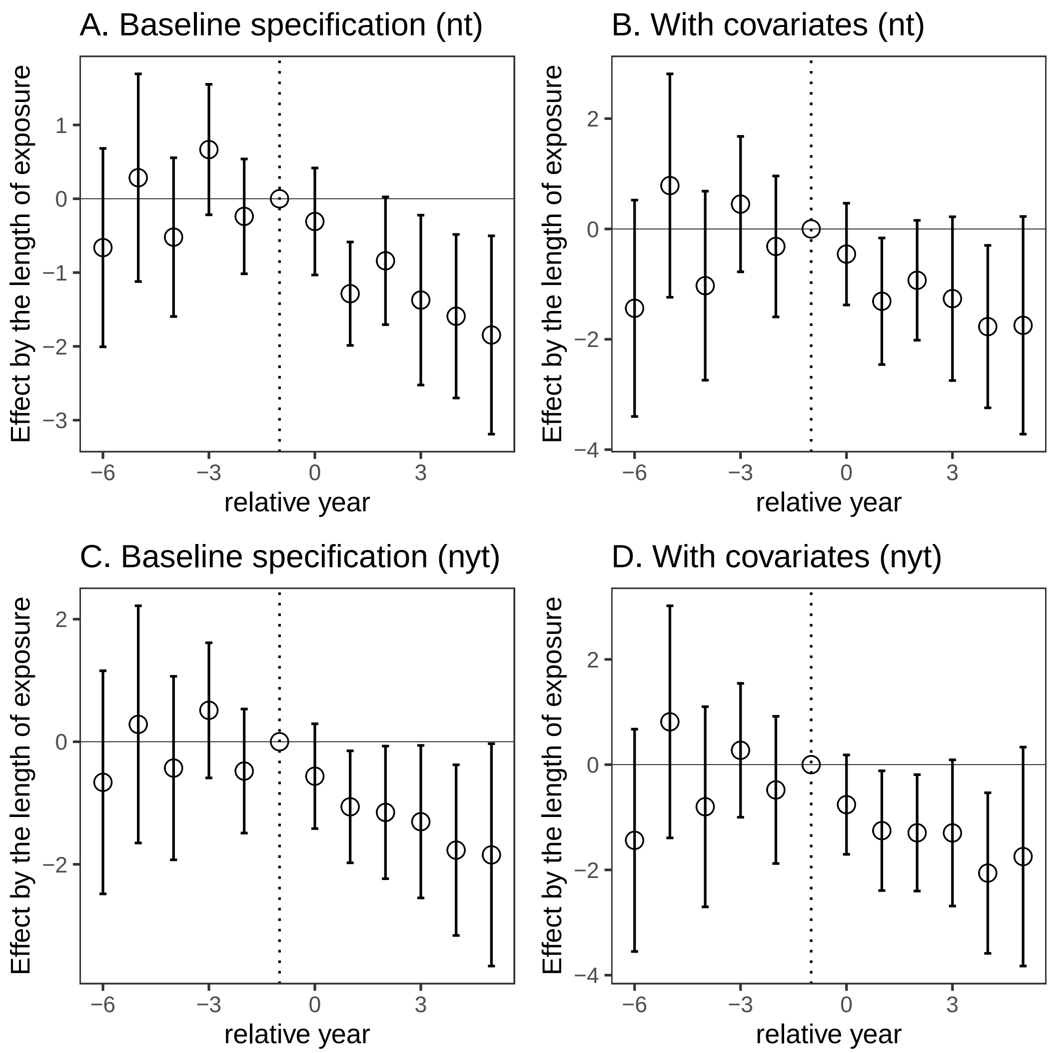

Figure 7: CS Event-Study-Type Estimates (never-treated and not-yet-treated as comparison)
Note: The figure shows the average treatment on the treated estimates by the length of exposure to the treatment using the CS estimator and 2004-2010 BRFSS SMART sample. Panels A and B use only the never-treated units as comparison, while Panels C and D use the not-yet-treated-units. Panels A and C present results without the covariates when the unconditional parallel trend assumption is invoked. The estimates in panels B and D use the DR approach and include the locality specific covariates: *i*) the log of population in 2000, *ii*) unemployment rate in 2000, *iii*) changes in the proportion of current smokers between 1998 and 2001, and *iv*) a measure of anti-smoking semtiment in 1998. The vertical bars represent 95% confidence intervals constructed using standard errors from multiplier bootstrapped procedure to account for multiple hypothesis testing.

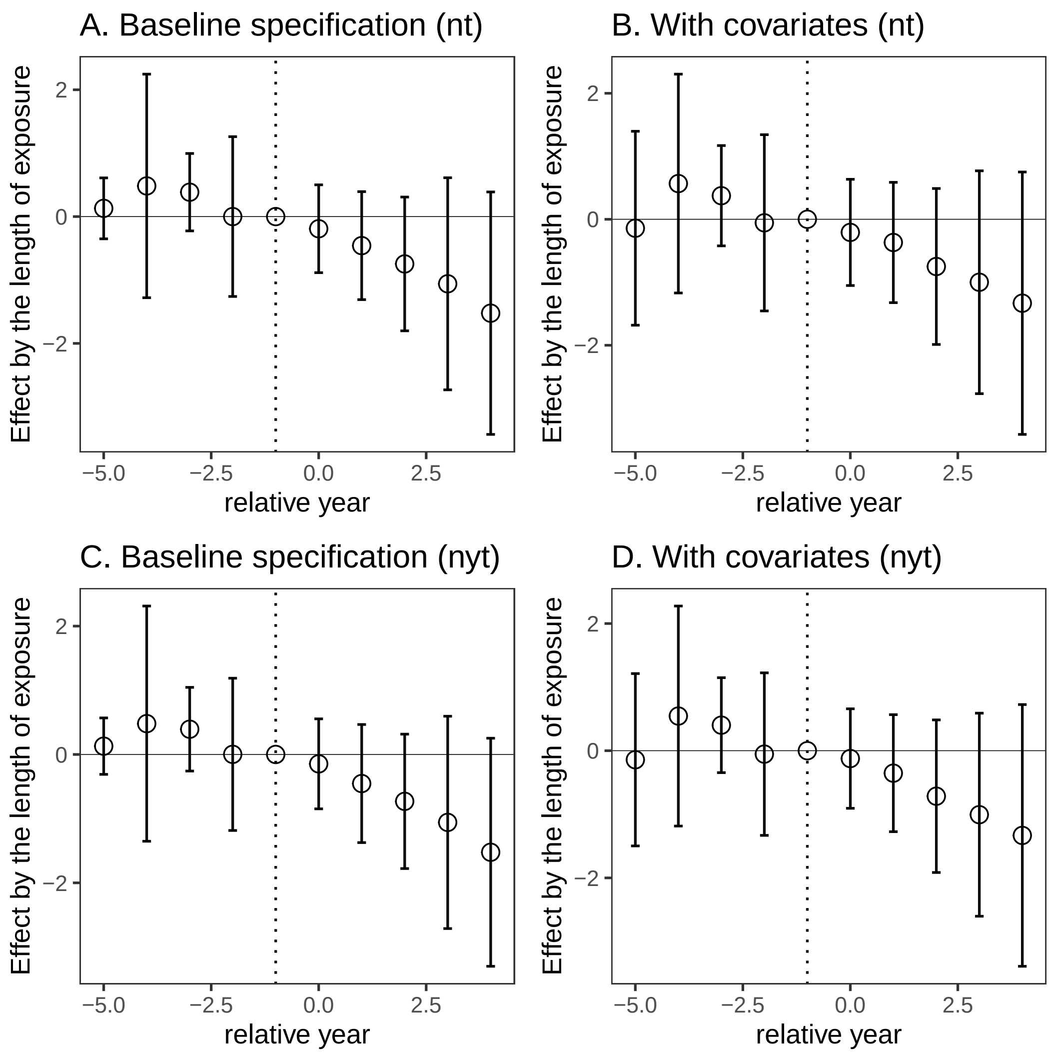

Figure 8: CS Event-Study-Type Estimates (never-treated and not-yet-treated as comparison) The figure shows the average treatment on the treated estimates by the length of exposure to the treatment using CS estimator and 2015-2020 BRFSS SMART sample. Panels A and B use only the never-treated units as comparison, while Panels C and D use the not-yet-treated-units. Panels A and C present results without the covariates when the unconditional parallel trend assumption is invoked. The estimates in panels B and D use the DR approach and include the locality specific covariates: *i*) the log of population in 2010, *ii*) unemployment rate in 2010, *iii*) changes in the proportion of current smokers between 1998 and 2001, and *iv*) a measure of anti-smoking semtiment in 1998. The vertical bars represent 95% confidence intervals constructed using standard errors from multiplier bootstrapped procedure to account for multiple hypothesis testing.

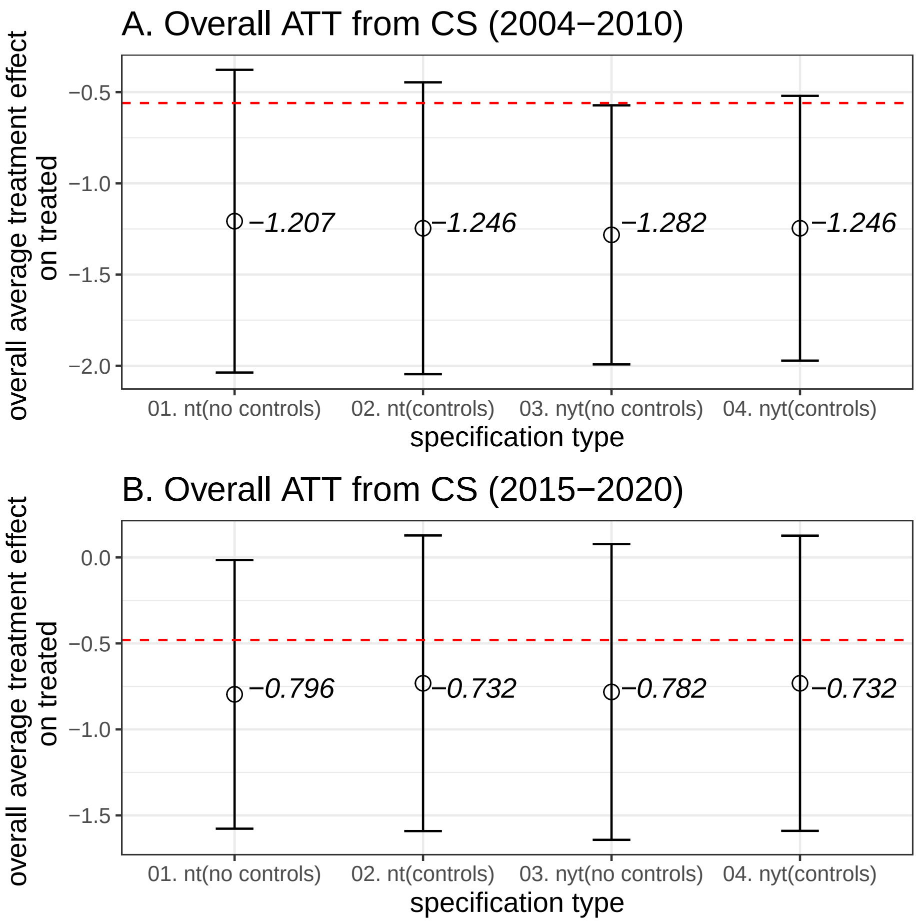

Figure 9: Overall ATT (using CS method)
The figures show the overall average treatment effect on the treated using the CS approach for the sample years 2004-2010 in Panel A and 2015-2020 in Panel B. Different versions of the CS estimator are considered. Approach 01 uses never-treated units as the comparison group without covariates. Approach 02 adds covariates and still uses never-treated units for comparison. Approaches 03 and 04 follow 01 and 02 but use not-yet-treated units to form the comparison groups. The locality-specific covariates included in 02 and 04 are: *i*) unemployment rate, *ii*) the log of population (in 2000 for panel A and 2010 for panel B), *iii*) a measure of anti-smoking sentiments in 1998, and *iv*) the change in the proportion of current smokers between 1998 and 2001. The vertical bar represents the 95% confidence interval.

Table 1: States with tax changes by year and MMAs count in BRFSS SMART

year	states	count of MMSAs	average tax increase ($)	s.d.
2004	AL, HW, MI, NJ, PA, RI, VA	11	0.34	0.2
2005	AK, CO, KY, ME, MN, MT, NC, NH, OH, OK, WA	28	0.5	0.31
2006	AZ, IA, VT	7	0.67	0.29
2007	CT, DE, IN, SD, TN, TX	19	0.75	0.24
2008	DC, MA, MD, NY, WI	9	1.01	0.17
2009	AR, FL, MS	7	0.74	0.23
2010	NM, SC, UT	9	0.75	0.21
–	–	–		
2015	DC, KS, LA, NV, OH, RI, VT	11	0.53	0.2
2016	AL, CT, PA, WV	7	0.65	0.33
2017	CA	1	2	NA
2018	DE, KY, OK	5	0.7	0.25
2019	IL, NM	3	0.78	0.32
2020	VA	1	0.3	NA

States implementing multiple tax changes within the sample period are identified by the earliest year of change. The count of MMSAs (Metropolitan and Micropolitan Statistical Areas) represents the number of MMSAs experiencing tax changes in the balanced BRFSS SMART sample.

Table 2: Summary statistics by units with and without a tax change (BRFSS SMART 2004-2010 & 2015-2020)

variables	tax change (04-10)	no change (04-10)	tax change (15-20)	no change (15-20)
prop. ever smoked	0.45	0.44	0.42	0.41
	(0.05)	(0.06)	(0.04)	(0.06)
% current smoker	19.57	21.03	14.74	13.51
	(4.04)	(4.53)	(2.81)	(2.89)
prop. stop smoking	0.59	0.57	0.59	0.58
	(0.05)	(0.06)	(0.06)	(0.06)
prop. current smoker (1998)	0.22	0.21	0.23	0.22
	(0.03)	(0.04)	(0.03)	(0.03)
nominal cig tax	1.36	0.71	1.9	1.79
	(0.78)	(0.39)	(1.01)	(1.16)
population 1990	5662925.39	5101805.57	5930206.89	6420580.91
	(4247435.7)	(5731874.82)	(5999860.9)	(5632039.15)
population 2000	6500855.7	5939002.2	6365436.98	7365499.56
	(4995460.36)	(6692914.24)	(6578373.17)	(6476565.66)
unemployment 1990	5.32	5.43	5.55	5.34
	(0.85)	(1.04)	(0.92)	(1.11)
unemployment 2000	3.67	3.9	4.04	3.68
	(0.84)	(0.82)	(0.89)	(0.83)
cig sales percapita (1990)	106.95	99.91	105.16	100.05
	(22.9)	(20.69)	(19.11)	(19.35)
cig sales percapita (2000)	86.46	80.59	90.94	80.33
	(28.64)	(23.99)	(22.75)	(23.25)
cig sales percapita (2010)	51.58	51.25	58.36	49.61
	(21.72)	(20.32)	(21.41)	(18.16)
smoking sentiment (1998)	-0.15	-0.09	-0.23	-0.13
	(0.16)	(0.2)	(0.15)	(0.18)
% under bar ban	35.47	26.59	62.7	75.14
	(37.11)	(33.99)	(45.9)	(39.55)

The table reports mean and standard deviation (in parenthesis) for the variables used in the study. Columns 2-3 pertain to MMSAs with and without the tax change in the BRFSS SMART 2004-2010 sample, while columns 4-5 refer to the 2015-2020 sample. Both BRFSS SMART 2004-2010 and 2015-2020 are balanced panels with 108 and 95 MMSAs. The percent of population living under the bar ban is calculated using 2005 and 2010 population size for the BRFSS SMART 2004-2010 and 2015-2020 samples, respectively.

Table 3: TWFE Estimates (BRFSS SMART 2004-2010 & 2015-2020)

	Type	weight (04-10)	avg. estimate (04-10)	weight (15-20)	avg. estimate (15-20)
1	Earlier vs Later Treated	0.218	-0.898	0.029	-0.048
2	Later vs Always Treated	0.177	-0.296	0.13	-0.471
3	Later vs Earlier Treated	0.316	-0.233	0.047	0.354
4	Treated vs Untreated	0.29	-0.834	0.794	-0.549

The table provides a summary of Goodman Bacon decomposition of the TWFE estimate as all possible 2×2 DiD estimates, summarized by groups mentioned in Column 1. Columns 2-3 and 4-5 refer to the sample years 2004-2010 and 2015-2020. The weight column corresponds to weights given to the respective group when using estimates to form the TWFE estimate. For instance, the sum of weight (04-10)$\times$ avg. estimate (04-10) equals to TWFE estimate in Column 1, Table 4 (Panel A).

Table 4: TWFE Estimates (BRFSS SMART 2004-2010 & 2015-2020)

	Panel A. % current smoker					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.5631**	-0.5612**	-0.6128***	-0.4824	-0.4592	-0.4105
	(0.2115)	(0.2113)	(0.2117)	(0.3196)	(0.3230)	(0.3423)
Observations	756	756	756	570	570	570
R^2	0.88960	0.88961	0.89378	0.86611	0.86629	0.87147

	Panel B. % current smoker (drop always treated)					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.6205***	-0.6058***	-0.6782***	-0.4840	-0.4675	-0.3438
	(0.2142)	(0.2148)	(0.2341)	(0.3260)	(0.3294)	(0.3471)
Observations	679	679	679	504	504	504
R^2	0.88855	0.88867	0.89270	0.87338	0.87346	0.87944
Year and MMSA FE	✓	✓	✓	✓	✓	✓
Smoke Free air laws		✓	✓		✓	✓
Additional Controls			✓			✓
BRFSS SMART 2004-10	✓	✓	✓			
BRFSS SMART 2014-20				✓	✓	✓

The table uses BRFSS MMSA SMART data for the years: *i*) 2004 to 2010 in Columns (1)-(3), and *ii*) 2015-2020 in Columns (4)-(6). All columns include the locality level (MMSA) and year fixed effects. Columns (1) and (4) depict results from the parsimonious specification that includes a dummy for whether a state increased excise taxes on cigarettes at time t. Once this indicator turns on, it remains on for the rest of the years in the panel. Columns (2) and (5) control for the percent of state population living under smoking ban in bars. Additionally, Columns (3) and (6) include a vector of pre-treatment variables (*i*. log of population measured in 2000 for the 2004-2010 sample and in 2010 for the 2015-2020 sample, *ii*. changes in the proportion of current smokers between 1998 and 2001, *iii*. the measure for anti-smoking sentiment in 1998, *iv*. unemployment rate in 2000 for 2004-2010 sample and in 2010 for the 2015-2020 sample) interacted with year dummies. Panels A and B are structured in a similar way, except that Panel B drops units that fall under always treated group. The standard errors clustered at the state level are presented in parenthesis.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

9 References

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* 72 (1): 1–19.

Adda, Jerome, and Francesca Cornaglia. 2006. "Taxes, Cigarette Consumption, and Smoking Intensity." *American Economic Review* 96 (4): 1013–28.

Anderson, D Mark, Kyutaro Matsuzawa, and Joseph J Sabia. 2020. "Cigarette Taxes and Teen Marijuana Use." *National Tax Journal* 73 (2): 475–510.

Barnett, Paul G, Theodore E Keeler, and Teh-wei Hu. 1995. "Oligopoly Structure and the Incidence of Cigarette Excise Taxes." *Journal of Public Economics* 57 (3): 457–70.

Baum, Charles L. 2009. "The Effects of Cigarette Costs on BMI and Obesity." *Health Economics* 18 (1): 3–19.

Bishop, James M. 2018. "Does Cigarette Smuggling Prop up Smoking Rates?" *American Journal of Health Economics* 4 (1): 80–104.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." *arXiv Preprint arXiv:2108.12419*.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment." *arXiv Preprint arXiv:2107.02637*.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Callison, Kevin, and Robert Kaestner. 2014. "Do Higher Tobacco Taxes Reduce Adult Smoking? New Evidence of the Effect of Recent Cigarette Tax Increases on Adult Smoking." *Economic Inquiry* 52 (1): 155–72.

Carpenter, Christopher, and Philip J Cook. 2008. "Cigarette Taxes and Youth Smoking: New Evidence from National, State, and Local Youth Risk Behavior Surveys." *Journal of Health Economics* 27 (2): 287–99.

Chaloupka, Frank J, and Kenneth E Warner. 2000. "The Economics of Smoking." *Handbook of Health Economics* 1: 1539–1627.

Conway, Karen Smith, and David P Niles. 2017. "Cigarette Taxes, Smoking—and Exercise?"

Health Economics 26 (8): 1019–36.

Cotti, Chad D, Charles J Courtemanche, Johanna Catherine Maclean, Erik T Nesson, Michael F Pesko, and Nathan Tefft. 2020. "The Effects of e-Cigarette Taxes on e-Cigarette Prices and Tobacco Product Sales: Evidence from Retail Panel Data." National Bureau of Economic Research.

Cotti, Chad, Erik Nesson, and Nathan Tefft. 2016. "The Effects of Tobacco Control Policies on Tobacco Products, Tar, and Nicotine Purchases Among Adults: Evidence from Household Panel Data." *American Economic Journal: Economic Policy* 8 (4): 103–23.

———. 2018. "The Relationship Between Cigarettes and Electronic Cigarettes: Evidence from Household Panel Data." *Journal of Health Economics* 61: 205–19.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–96.

DeCicca, Philip, Donald S Kenkel, and Michael F Lovenheim. 2020. "The Economics of Tobacco Regulation: A Comprehensive Review."

DeCicca, Philip, Donald Kenkel, Alan Mathios, Yoon-Jeong Shin, and Jae-Young Lim. 2008. "Youth Smoking, Cigarette Prices, and Anti-Smoking Sentiment." *Health Economics* 17 (6): 733–49.

Decker, Sandra, and Amy Ellen Schwartz. 2000. "Cigarettes and Alcohol: Substitutes or Complements?" National bureau of economic research Cambridge, Mass., USA.

Farrelly, Matthew C, Jeremy W Bray, Gary A Zarkin, and Brett W Wendling. 2001. "The Joint Demand for Cigarettes and Marijuana: Evidence from the National Household Surveys on Drug Abuse." *Journal of Health Economics* 20 (1): 51–68.

Farrelly, Matthew C, Terry F Pechacek, and Frank J Chaloupka. 2003. "The Impact of Tobacco Control Program Expenditures on Aggregate Cigarette Sales: 1981–2000." *Journal of Health Economics* 22 (5): 843–59.

Farrelly, Matthew C, Terry F Pechacek, Kristin Y Thomas, and David Nelson. 2008. "The Impact of Tobacco Control Programs on Adult Smoking." *American Journal of Public Health* 98 (2): 304–9.

Friedson, Andrew, Moyan Li, Katherine Meckel, Daniel I Rees, and Daniel W Sacks. 2023. "Cigarette Taxes, Smoking, and Health in the Long Run." *Journal of Public Economics* 222: 104877.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.

Gruber, Jonathan, and Michael Frakes. 2006. "Does Falling Smoking Lead to Rising Obesity?" *Journal of Health Economics* 25 (2): 183–97.

Hansen, Benjamin, Joseph J Sabia, and Daniel I Rees. 2017. "Have Cigarette Taxes Lost Their Bite? New Estimates of the Relationship Between Cigarette Taxes and Youth Smoking." *American Journal of Health Economics* 3 (1): 60–75.

Harding, Matthew, Ephraim Leibtag, and Michael F Lovenheim. 2012. "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data." *American Economic Journal: Economic Policy* 4 (4): 169–98.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64 (4): 605–54.

Hoehn-Velasco, Lauren, Michael F Pesko, and Serena Phillips. 2023. "The Long-Term Impact of in Utero Cigarette Taxes on Adult Prenatal Smoking." *American Journal of Health Economics* 9 (4): 605–48.

Jones, Andrew M. 1989. "A Double-Hurdle Model of Cigarette Consumption." *Journal of Applied Econometrics* 4 (1): 23–39.

Lewit, Eugene M, and Douglas Coate. 1982. "The Potential for Using Excise Taxes to Reduce Smoking." *Journal of Health Economics* 1 (2): 121–45.

Markowitz, Sara. 2008. "The Effectiveness of Cigarette Regulations in Reducing Cases of Sudden Infant Death Syndrome." *Journal of Health Economics* 27 (1): 106–33.

Markowitz, Sara, E Kathleen Adams, Patricia M Dietz, Van T Tong, and Viji Kannan. 2013. "Tobacco Control Policies, Birth Outcomes, and Maternal Human Capital." *Journal of Human Capital* 7 (2): 130–60.

Meyer, Breed D. 1995. “Natural and Quasi-Experiments in Economics.” *Journal of Business & Economic Statistics* 13 (2): 151–61.

Mullahy, John. 1985. *Cigarette Smoking: Habits, Health Concerns, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand.* University of Virginia.

Nesson, Erik. 2017. “Heterogeneity in Smokers’ Responses to Tobacco Control Policies.” *Health Economics* 26 (2): 206–25.

Patrick, Stephen W, Kenneth E Warner, Elisabeth Pordes, and Matthew M Davis. 2016. “Cigarette Tax Increase and Infant Mortality.” *Pediatrics* 137 (1).

Pesko, Michael F, Charles J Courtemanche, and Johanna Catherine Maclean. 2020. “The Effects of Traditional Cigarette and e-Cigarette Tax Rates on Adult Tobacco Product Use.” *Journal of Risk and Uncertainty* 60 (3): 229–58.

Picone, Gabriel A, Frank Sloan, and Justin G Trogdon. 2004. “The Effect of the Tobacco Settlement and Smoking Bans on Alcohol Consumption.” *Health Economics* 13 (10): 1063–80.

Rosenbaum, Paul R. 1984. “The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment.” *Journal of the Royal Statistical Society: Series A (General)* 147 (5): 656–66.

Roth, Jonathan, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe. 2022. “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.” *arXiv Preprint arXiv:2201.01194*.

Sant’Anna, Pedro HC, and Jun Zhao. 2020. “Doubly Robust Difference-in-Differences Estimators.” *Journal of Econometrics* 219 (1): 101–22.

Seldon, Barry J, and Roy Boyd. 1991. “The Stability of Cigarette Demand.” *Applied Economics* 23 (2): 319–26.

Shrestha, Vinish. 2018. “Do Young Adults Substitute Cigarettes for Alcohol? Learning from the Master Settlement Agreement.” *Review of Economics of the Household* 16: 297–321.

Sun, Liyang, and Sarah Abraham. 2021. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics* 225 (2): 175–99.

Sung, Hai-Yen, Teh-Wei Hu, and Theodore E Keeler. 1994. “Cigarette Taxation and Demand: An

Empirical Model." *Contemporary Economic Policy* 12 (3): 91–100.

Tauras, John, Lisa Powell, Frank Chaloupka, and Hana Ross. 2007. "The Demand for Smokeless Tobacco Among Male High School Students in the United States: The Impact of Taxes, Prices and Policies." *Applied Economics* 39 (1): 31–41.

Wasserman, Jeffrey, Willard G Manning, Joseph P Newhouse, and John D Winkler. 1991. "The Effects of Excise Taxes and Regulations on Cigarette Smoking." *Journal of Health Economics* 10 (1): 43–64.

Williams, Jenny, Rosalie Liccardo Pacula, Frank J Chaloupka, and Henry Wechsler. 2004. "Alcohol and Marijuana Use Among College Students: Economic Complements or Substitutes?" *Health Economics* 13 (9): 825–43.

10 Appendix

01. Current Population Survey: Tobacco Supplement 1998-1999 and 2001-2000

Not all of the MMSAs from BRFSS SMART data are represented in CPS tobacco supplement files. In fact 39 MMSAs in 2004-2010 BRFSS SMART balanced panel data cannot be mapped to MSA/PMSA codes in CPS files. In light of using all observations in the balanced panel, I impute the missing values on: *i*) the measure of anti-smoking sentiments in 1998, and *ii*) the proportion of current smokers in 1998 and 2001. First, the non-missing values on the outcomes are regressed on the locality specific wage index and the log of population as well as state specific variables including the log of cigarette sales in 2000. Next, the estimates from the regression are then use to construct predicted values of the outcomes. The missing values on the outcomes are replaced by the predicted values. Figure 10 in the Appendix plots the relationship between the measure of anti-smoking sentiments and smoking outcomes.

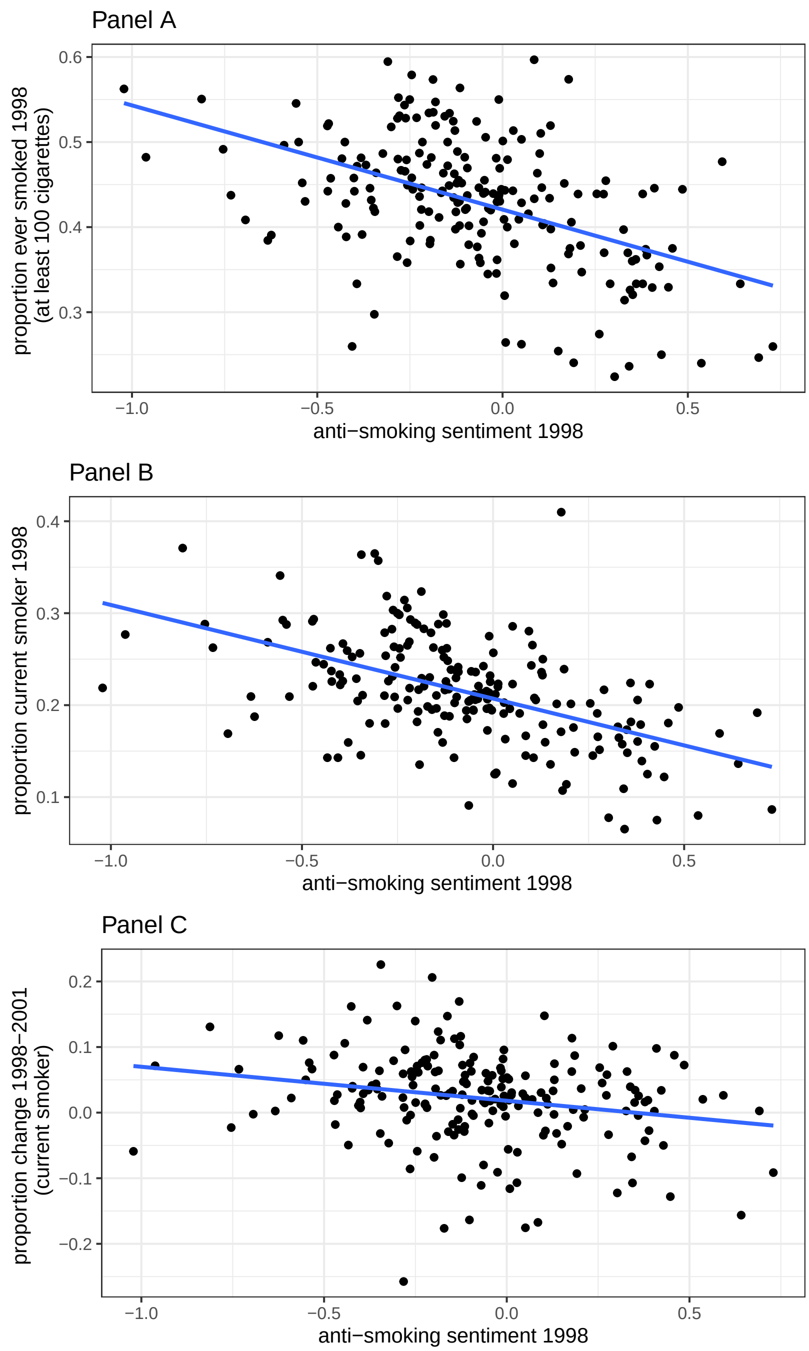

Figure 10: Anti-smoking sentiments and smoking
Note: The sub-figures show the relationship between the measure of anti-smoking sentiments in 1998 constructed by using the principal factor analysis and smoking-related variables. Panel A uses the propotion of individuals having reported ever smoked a cigarette (100 cigarettes or more), Panel B uses the proportion of current smokers, and Panel C uses the change in the proportion of current smokers between 1998 and 2001. The data comes from the Current Population Survey (Tobacco Supplement) files and is aggregated at the MMSA level.

02. Other Results

A. Evolution of smoking ban in bars

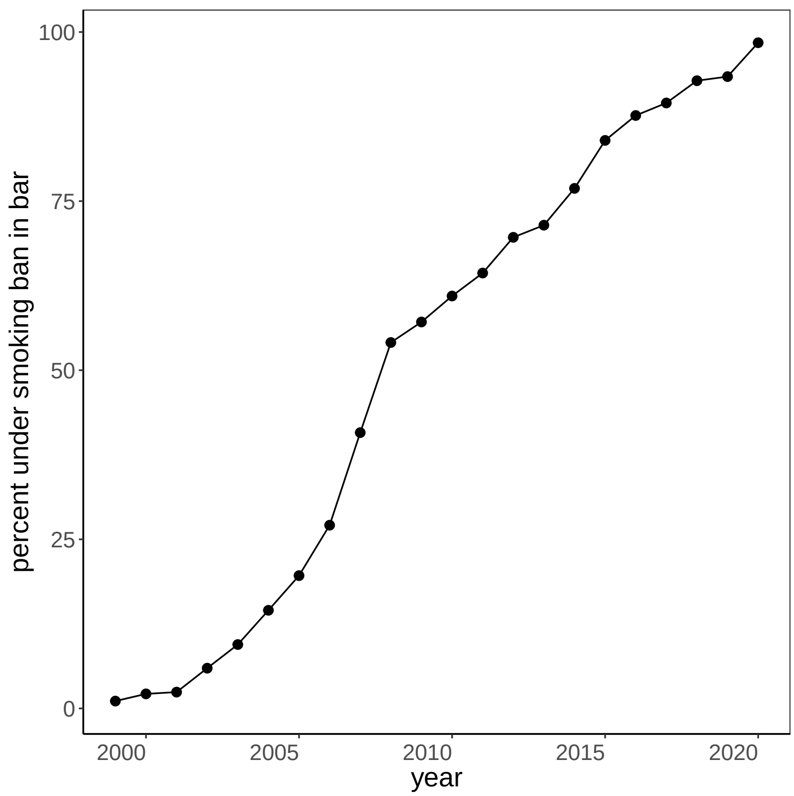

Figure 11: Percent of population under bar ban
Note: The figure illustrates the evolution of the percentage of the population living under smoking bans in bars from the year 2000 onwards. The data for smoking ban in bars is collected at the local level and sourced from the American Nonsmokers' Rights Foundation (ANRF). Population estimates from the Core-Based Statistical Area (CBSA) are utilized to calculate the percentage of the population living under the bar ban policy for the years 2000 to 2020, which are then aggregated at the state level. These estimates are merged with BRFSS SMART survey years 2000-2010 and 2015-2020.

B. Dosage of increases in cigarette taxes and pre-treatment variables

Table 5: Cigarette tax dosage and pre-treatement variables

	Dependent variable:							
	cig sales 1990	cig sales 2000	anti smoking	bar ban	cig sales 1990	cig sales 2000	anti smoking	bar ban
	2004-2010				2015-2020			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
tax change dose	-45.76^{***}	-49.36^{***}	0.26^{***}	7.00	-15.75	-17.92	0.04	-35.35
	(6.22)	(8.10)	(0.06)	(9.58)	(10.77)	(12.90)	(0.09)	(22.27)
Observations	90	90	90	90	28	28	28	28

Note: *p<0.1; **p<0.05; ***p<0.01

Columns 1-4 and 5-8 refer to 2004-2010 and 2015-2020 samples. The specifications are conditional on units experiencing increases in cigarette tax within the sample period. The tax change dose represents the nominal increase in cigarette taxes. Columns 1-2 and 5-6 use the per capita cigarette sales for 1990 and 2000 as the dependent variable. Columns 3 and 7 use the measure of anti-smoking sentiment in 1998, while Columns 4 and 8 use the percent of population living under bar ban in 2010.

Table 5 shows that there are large differences in some important pre-treatment variables across different intensity of cigarette tax increases (dosage) among units experiencing tax increases between 2004-2010. For example, levels of tax increases are negatively correlated with pre-treatment per capita cigarette sales. Moreover, units with higher levels of anti-smoking sentiment measure are more likely to implement larger tax hikes. Hence, units facing higher vs. lower levels of taxes between 2004-2010 tend to be very different in pre-treatment variables that may also have systematically affected the evolution of smoking outcomes even in absence of tax increases.

The identification when using a continuous measure of cigarette taxes depends on not just the parallel trends in outcome between the treated and untreated groups but also parallel trends across units receiving various dosage of treatments.[29] For example, consider the case when unit A receives a dose of d and unit B receives a dose of d', where $d > d'$. In this case, the validity of identification depends on the assumption that unit A's outcome would have evolved similarly to unit B had unit A received dosage d' instead of d. This assumption is stronger than the parallel trend assumption based on the binary treatment measure as the potential issue of selection when using a continuous

[29] Comparison of outcome changes between units receiving the treatment of dose d with untreated units gives the average treatment effect of dose d. Comparison of outcomes changes between units receiving two different dosage can be considered as causal responses. See Callaway, Goodman-Bacon, and Sant'Anna (2021) for detailed discussions.

treatment measure is not just in terms of who gets treated but also the levels of treatment.

C. Controling for tax implementation later in the year

Table 6: TWFE Estimates (BRFSS SMART 2004-2010 & 2015-2020)

	Panel A. % current smoker					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.5826***	-0.5812***	-0.6340***	-0.5218	-0.4985	-0.4494
	(0.2127)	(0.2133)	(0.2164)	(0.3354)	(0.3389)	(0.3589)
Observations	756	756	756	566	566	566
R^2	0.88966	0.88966	0.89383	0.86592	0.86608	0.87123
	Panel B. % current smoker (drop always treated)					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.6372***	-0.6212***	-0.7041***	-0.5193	-0.5029	-0.3765
	(0.2174)	(0.2183)	(0.2440)	(0.3418)	(0.3453)	(0.3641)
Observations	679	679	679	501	501	501
R^2	0.88859	0.88870	0.89277	0.87320	0.87327	0.87918
Year and MMSA FE	✓	✓	✓	✓	✓	✓
Smoke Free air laws		✓	✓		✓	✓
Additional Controls			✓			✓
BRFSS SMART 2004-10	✓	✓	✓			
BRFSS SMART 2014-20				✓	✓	✓

The table uses BRFSS MMSA SMART data for years: *i)* 2004 to 2010 in Columns (1)-(3), and *ii)* 2015-2020 in Columns (4)-(6). All specifications include control for an indicator representing whether the tax implementation was during the later part of the year (past July) in the year of implementation. All columns include the locality level (MMSA) and year fixed effects. Columns (1) and (4) depict results from the parsimonious specification that includes a dummy for whether a state increased excise taxes on cigarettes at time t. Once this indicator turns on, it remains on for the rest of the years in the panel. Columns (2) and (5) control for the percent of state population living under smoking ban in bars. Additionally, Columns (3) and (6) include a vector of pre-treatment variables (*i.* log of population measured in 2000 for the 2004-2010 sample and in 2010 for the 2015-2020 sample, *ii.* changes in the proportion of current smokers between 1998 and 2001, *iii.* the measure for anti-smoking sentiment in 1998, *iv.* unemployment rate in 2000 for 2004-2010 sample and in 2010 for 2015-2020 sample) interacted with year dummies. Panels A and B are structured in a similar way, except that Panel B drops units that fall under always treated group. The standard errors clustered at the state level are presented in parenthesis. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

10.1 D. Considering the implementation year as the adjustment period

Table 7: TWFE Estimates (BRFSS SMART 2004-2010 & 2015-2020)

	Panel A. % current smoker					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.5846***	-0.5814***	-0.6365***	-0.5218	-0.4985	-0.4460
	(0.2131)	(0.2137)	(0.2149)	(0.3351)	(0.3386)	(0.3600)
Observations	745	745	745	563	563	563
R^2	0.89021	0.89022	0.89445	0.86579	0.86595	0.87115
	Panel B. % current smoker (drop always treated)					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for tax increase	-0.6372***	-0.6208***	-0.6997***	-0.5193	-0.5029	-0.3721
	(0.2173)	(0.2181)	(0.2414)	(0.3415)	(0.3450)	(0.3657)
Observations	669	669	669	498	498	498
R^2	0.88923	0.88935	0.89329	0.87307	0.87315	0.87912
Year and MMSA FE	✓	✓	✓	✓	✓	✓
Smoke Free air laws		✓	✓		✓	✓
Additional Controls			✓			✓
BRFSS SMART 2004-10	✓	✓	✓			
BRFSS SMART 2014-20				✓	✓	✓

The table uses BRFSS MMSA SMART data for years: *i)* 2004 to 2010 in Columns (1)-(3), and *ii)* 2015-2020 in Columns (4)-(6). The first year of tax hike is considered as the adjustment year regardless of the month of implementation. As such, in this setup the treatment occurs only a year after the implementation year. All columns include the locality level (MMSA) and year fixed effects. Columns (1) and (4) depict results from the parsimonious specification that includes a dummy for whether a state increased excise taxes on cigarettes at time t. Once this indicator turns on, it remains on for the rest of the years in the panel. Columns (2) and (5) control for the percent of state population living under smoking ban in bars. Additionally, Columns (3) and (6) include a vector of pre-treatment variables (*i.* log of population measured in 2000 for the 2004-2010 sample and in 2010 for the 2015-2020 sample, *ii.* changes in the proportion of current smokers between 1998 and 2001, *iii.* the measure for anti-smoking sentiment in 1998, *iv.* unemployment rate in 2000 for 2004-2010 sample and in 2010 for 2015-2020 sample) interacted with year dummies. Panels A and B are structured in a similar way, except that Panel B drops units that fall under always treated group. The standard errors clustered at the state level are presented in parenthesis. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

10.2 E. Canonical event-study estimates

Table 8: Event-Study Estimates (BRFSS SMART 2004-2010)

	% current smoker	
	(1)	(2)
treat × year_around = -5	0.2139	0.1086
	(0.3693)	(0.4710)
treat × year_around = -4	-0.2599	-0.3627
	(0.3914)	(0.4246)
treat × year_around = -3	0.2786	0.3376
	(0.2902)	(0.2981)
treat × year_around = -2	-0.4320	-0.4039
	(0.3065)	(0.2982)
treat × year_around = 0	-0.4969*	-0.5593*
	(0.2586)	(0.2817)
treat × year_around = 1	-1.186***	-1.231***
	(0.3479)	(0.3528)
treat × year_around = 2	-1.153***	-1.175***
	(0.3610)	(0.3618)
treat × year_around = 3	-1.533***	-1.518***
	(0.4366)	(0.4347)
treat × year_around = 4	-1.415***	-1.430***
	(0.4913)	(0.4695)
treat × year_around = 5	-1.614***	-1.654***
	(0.5394)	(0.5404)
Observations	679	679
R^2	0.89179	0.89564
Year and MMSA FE	✓	✓
Smoke Free air laws		✓
Additional Controls		✓
BRFSS SMART 2004-10		✓
BRFSS SMART 2014-20		✓

The table uses BRFSS MMSA SMART data for the years 2004 to 2010. The always-treated units are excluded. All columns include the locality level (MMSA) and year fixed effects. Columns (2) include a vector of pre-treatment variables (*i*. log of population measured in 2000, *ii*. changes in the proportion of current smokers between 1998 and 2001, *iii*. the measure for anti-smoking sentiment in 1998, *iv*. unemployment rate in 2000) interacted with year dummies. The standard errors clustered at the state level are presented in parenthesis. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 9: Event-Study Estimates (BRFSS SMART 2015-2020)

	% current smoker	
	(1)	(2)
treat × year_around = -4	0.2639	0.3368
	(0.5847)	(0.6206)
treat × year_around = -3	0.3086	0.2186
	(0.3525)	(0.3764)
treat × year_around = -2	-0.0920	-0.1685
	(0.5934)	(0.5432)
treat × year_around = 0	-0.1562	-0.1563
	(0.2535)	(0.3183)
treat × year_around = 1	-0.4757	-0.3980
	(0.4387)	(0.4390)
treat × year_around = 2	-0.6444	-0.6067
	(0.4175)	(0.4850)
treat × year_around = 3	-0.9937	-0.8983
	(0.6931)	(0.6429)
treat × year_around = 4	-1.306**	-1.060*
	(0.5453)	(0.6279)
Observations	504	504
R^2	0.87504	0.88036
Year and MMSA FE	✓	✓
Smoke Free air laws		✓
Additional Controls		✓
BRFSS SMART 2004-10		✓
BRFSS SMART 2014-20		✓

The table uses BRFSS MMSA SMART data for the years 2015 to 2020. The always-treated units are excluded. All columns include the locality level (MMSA) and year fixed effects. Columns (2) include a vector of pre-treatment variables (*i*. log of population measured in 2010, *ii*. changes in the proportion of current smokers between 1998 and 2001, *iii*. the measure for anti-smoking sentiment in 1998, *iv*. unemployment rate in 2010) interacted with year dummies. The standard errors clustered at the state level are presented in parenthesis. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

03. Simulation Setup

The simulated dataset is constructed by using the skeleton of the actual dataset, which includes 46 states and n number of MMSAs (n = 108 and 95 for the 2004-2010 and 2015-2020 samples, respectively).[30] The treatment assignment for units are similar to the actual treatment timing as demonstrated in Table 1. Let G_i define the group of unit i such that $G_i \subseteq \{2004,\ 2005, ..., 2010\}$ for the 2005-2010 sample, while g represents the first period of the treatment. The data generating process (DGP) is written as:

$$Y_{it} = \tau_{it} + \eta_i + \theta_t + \epsilon_{it} \tag{10}$$

where, τ_{it} is the treatment effect, η_i is the unit fixed effects, and θ_t is the time fixed effects. The unit fixed effects (η_i) is drawn from $N \sim (0, 20)$ and is specific to each MSA. The time fixed effects (common across units) are generated as:

$$\theta_t = \rho \times (t - 2003) + v_t \tag{11}$$

where, ρ is set equal to 1 and $v_t \sim N(0,\ 1)$.

The treatment effects captured by τ_{it} is given as:

$$\tau_{it} = \sum_{k=-6}^{5} 1(t - g = k) \times \tau_k + \tau_i \tag{12}$$

Here, k is the relative time of treatment. $1(t - g = k)$ is an indicator representing whether the calendar period t defines the relative time k based on the initiation of treatment (g) for unit i. τ_k is the treatment effect specific to the relative time k. Additionally, τ_i represents the unit-specific treatment effect for the unit i in group G. The following three conditions define the categories of

[30] Although the results discussed here specifically pertains to the 2004-2005 sample, the simulation approach for the later sample is similar.

treatment effects:

i) Homogeneous treatment effect: $\tau_k = 0$ and $\tau_i = \tau$ for all i. $\tau_{it} = \tau$, τ is set to -1.2.

ii) Heterogeneous treatment effects across groups: $\tau_k = 0$ and τ_i varies across group.

iii) Heterogeneous treatment effects by relative time: $\tau_i = 0$ but τ_k varies across relative time.

The true treatment effects for case *ii*) are set as below:

group	ATT
2005	-1.396
2006	-2.685
2007	0.226
2008	-0.794
2009	-0.719
2010	-0.661

The true treatment effects for case *iii*) are set as:

relative.time	ATT
0	-0.3077
1	-1.2863
2	-0.8405
3	-1.3728
4	-1.5916
5	-1.8456

I perform 5,000 replication of each simulations for case *i*), *ii*) and *iii*) and evaluate the performance of the TWFE and CS estimator. The findings in Figures 12, 13, and 14 in Appendix show that the CS estimator performs well in all three cases.

04. Simulation Results

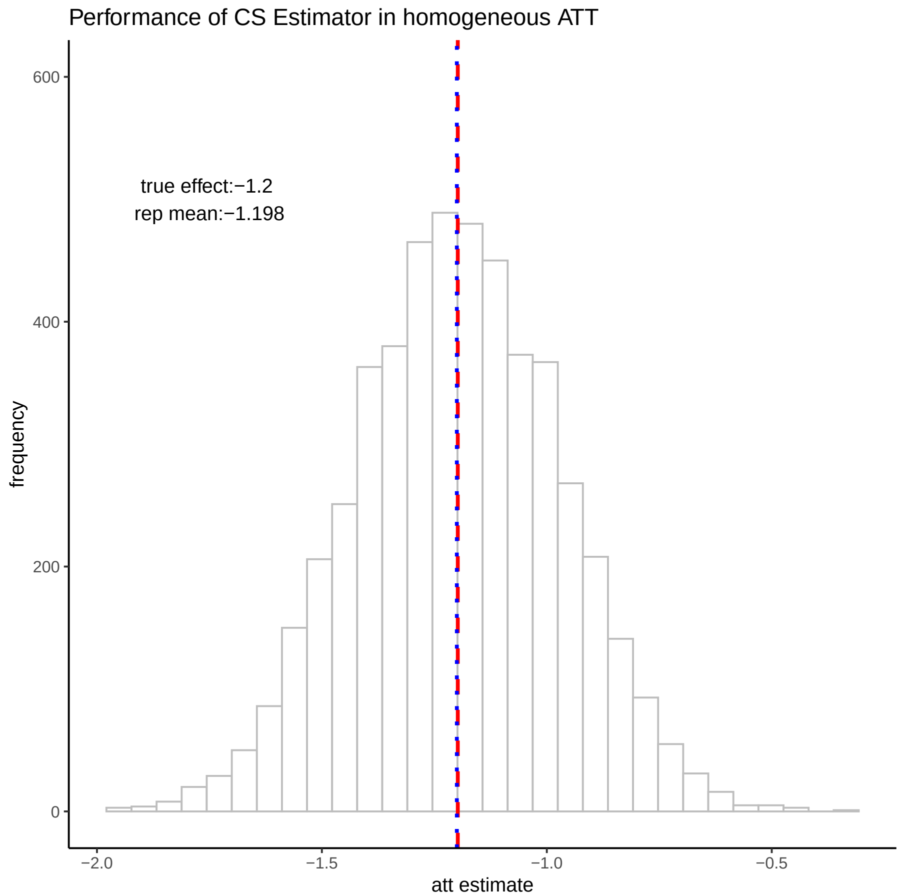

Figure 12: Simulation – CS Estimator

Note: The figure presents results from the simulation exercise that uses the variation in tax changes for the sample period 2004-2010 and pertains to the case of homogeneous treatment effects. The simulation consist of 5,000 rounds of replications. The true ATT effect and the mean estimates from replications are noted on the top-left segment with years following the treatment listed on the x-axis.

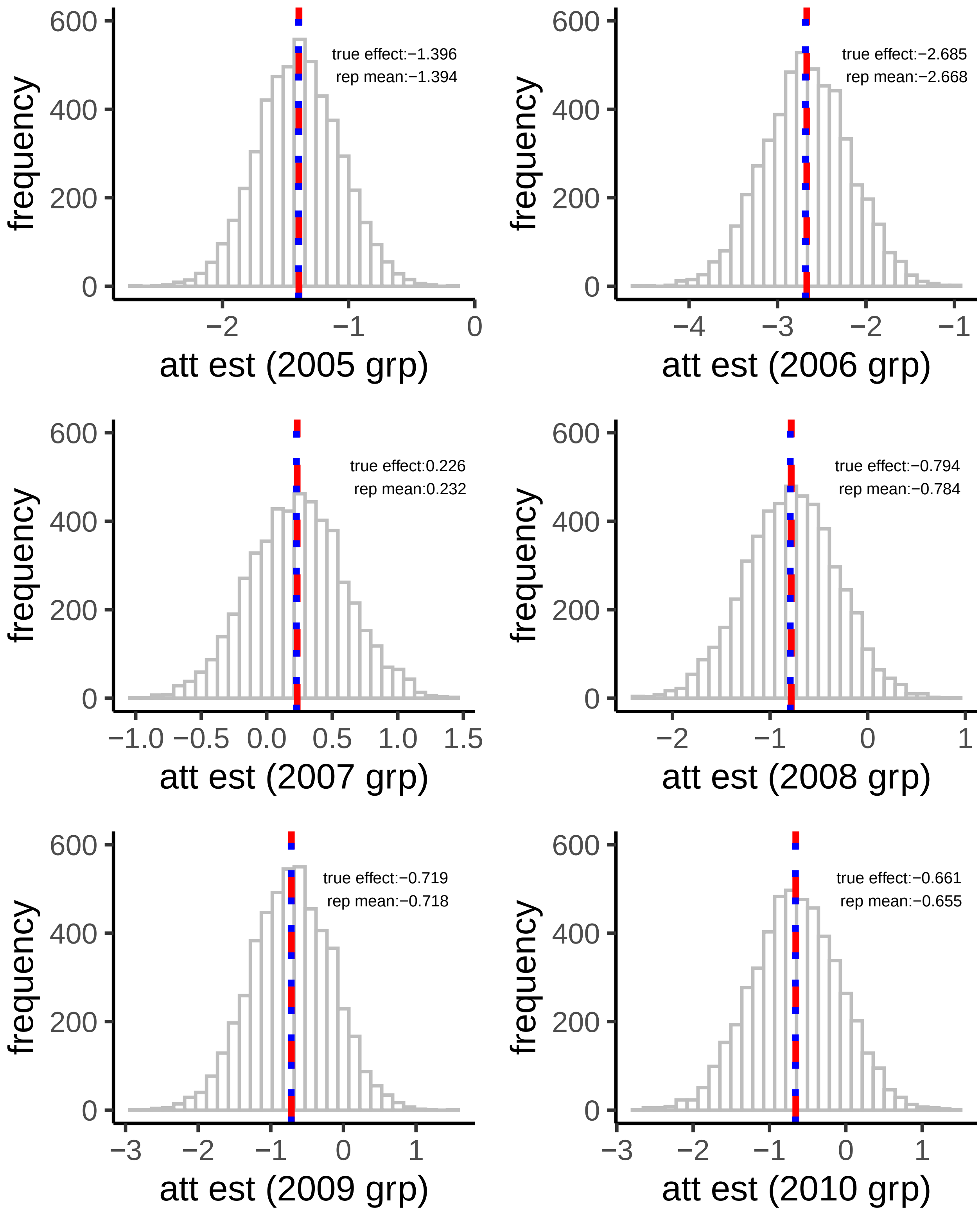

Figure 13: Simulation – CS Estimator
Note: The figure presents results from the simulation exercise that uses the variation in tax changes for the sample period 2004-2010 and pertains to the case of heterogeneous treatment effects by group. The simulation consist of 5,000 rounds of replications. The true ATT effect and the mean estimates from replications are noted on the top-left segment with years following the treatment listed on the x-axis.

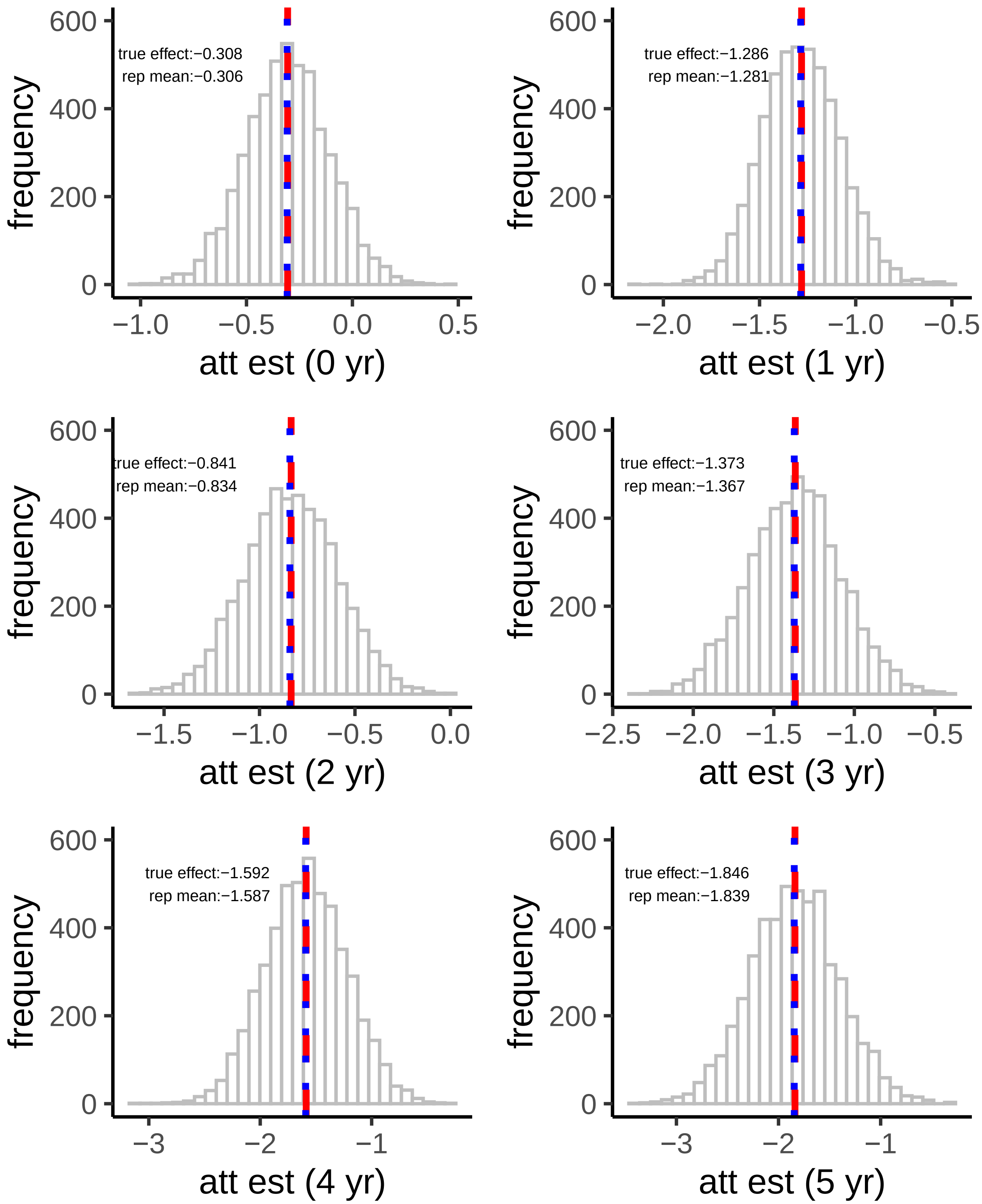

Figure 14: Simulation – CS Estimator
Note: The figure presents results from the simulation exercise that uses the variation in tax changes for the sample period 2004-2010 and pertains to the case of heterogeneous treatment effects by relative time. The simulation consist of 5,000 rounds of replications. The true ATT effect and the mean estimates from replications are noted on the top-left segment with years following the treatment listed on the x-axis.

04. CS Estimates (balanced in relative time)

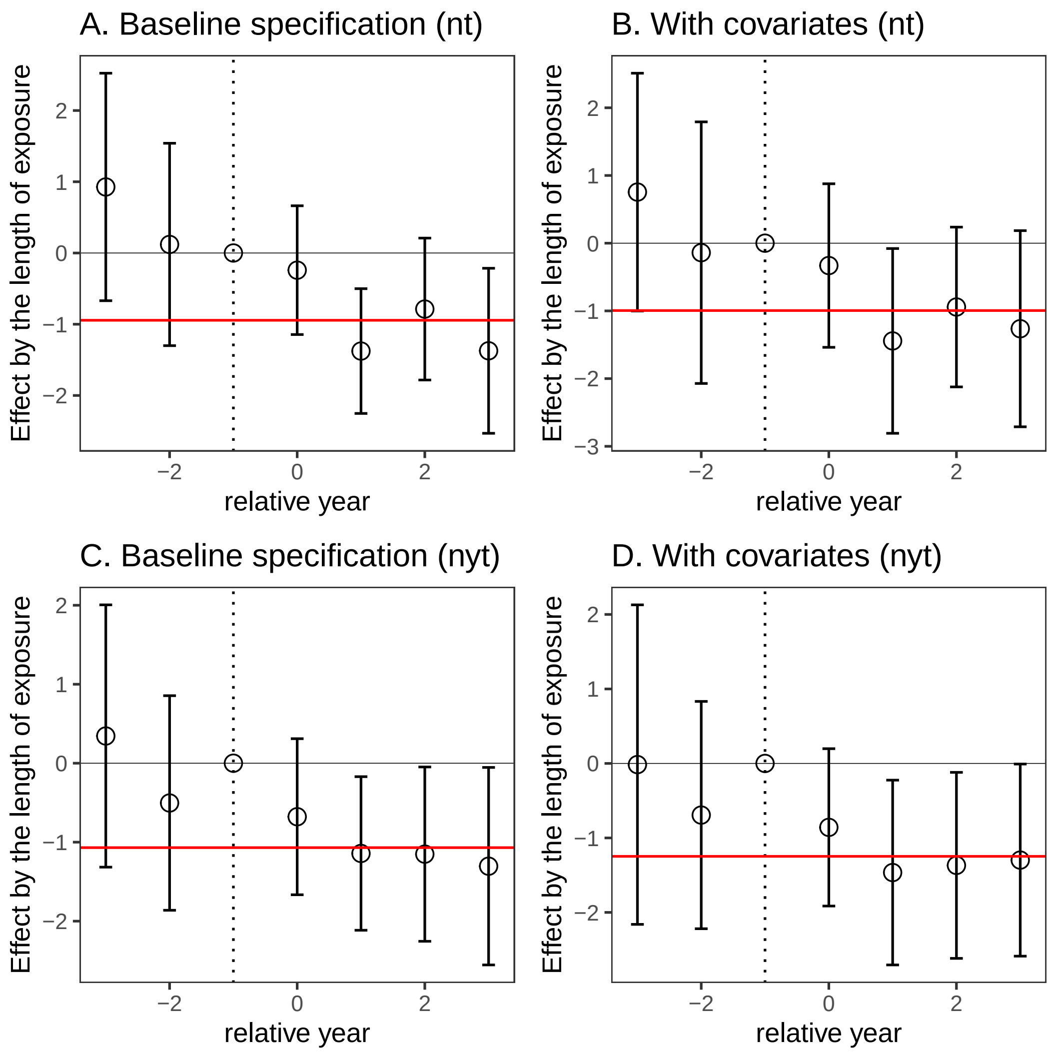

Figure 15: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison) Note: The figure is structured similar to Figure 5 except that the estimation only consist of units that are exposed to the treatment for at least 3 years following the treatment year. The red line represents the point estimate of the overall average treatment effect on the treated using the CS approach.

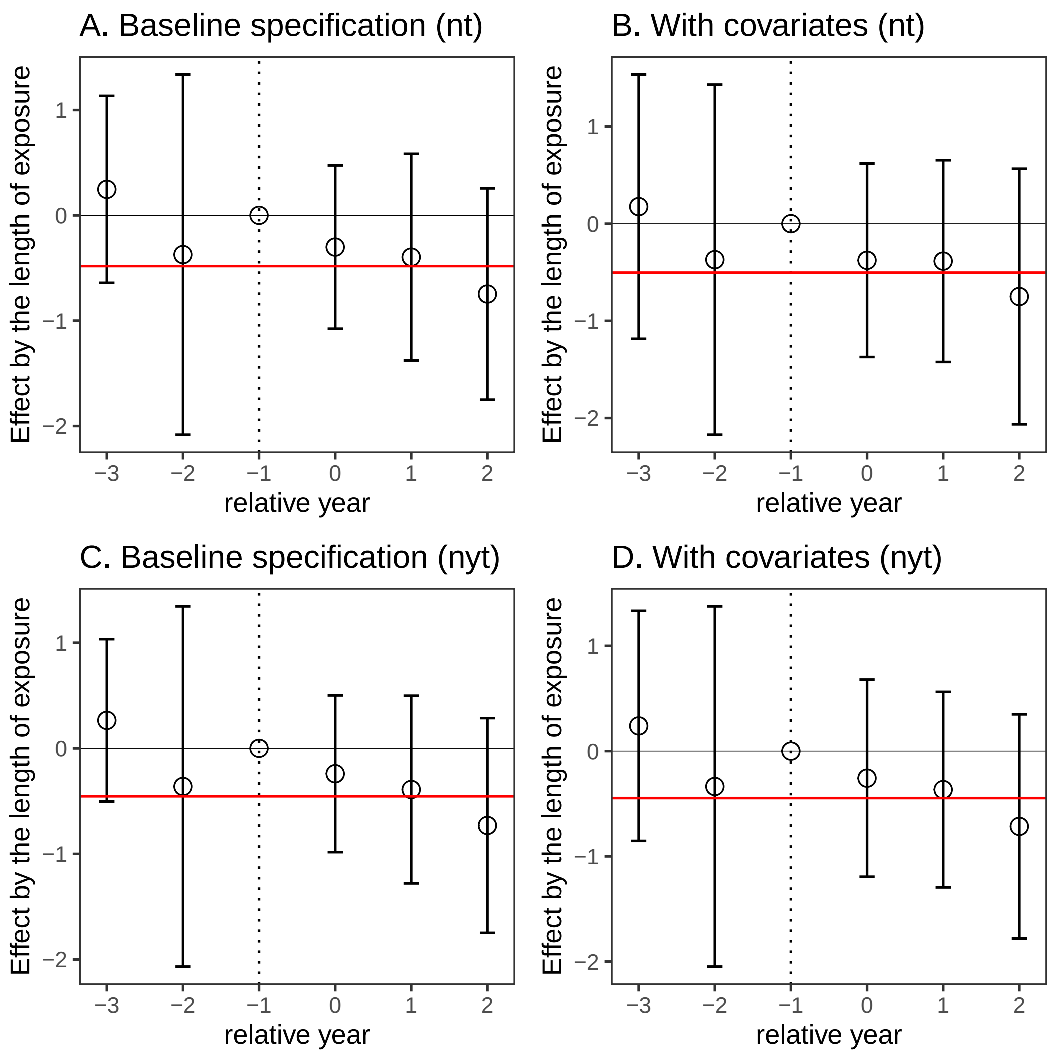

Figure 16: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison) The figure is structured similar to Figure 6 except that the panels only consist of units that are exposed to the treatment for at least 2 years following the treatment year. The red line represents the point estimate of the overall average treatment effect on the treated using the CS approach.

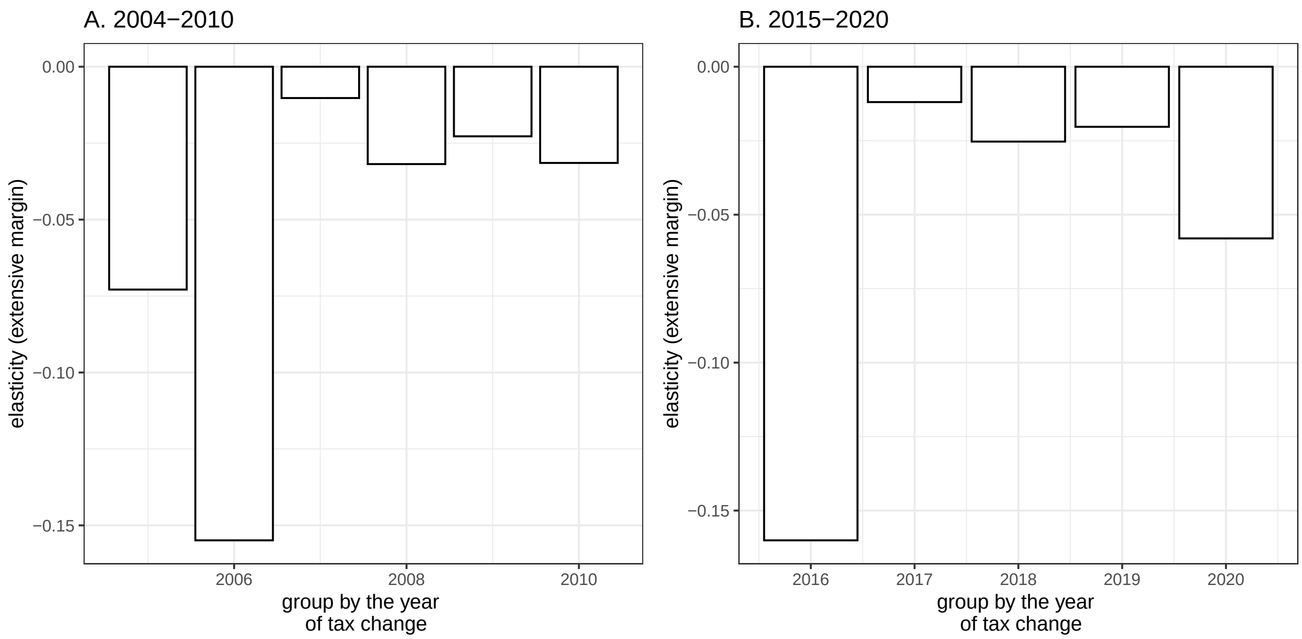

Figure 17: Elasticity Estimates (extensive margin)
The figures show the elasticity estimates when using current smokers as the dependent variable. The estimates should be interpreted as the results at the extensive margin defined as the percent reduction in current smokers due to a percent increase in cigarette tax. Panels A and B refers to the 2004-2010 and 2015-2020 samples, respectively.