

Does Quitting Smoking Increase Obesity? Evidence From Accounting for Misreporting

Rusty Tchernis*

Keith Teltser[†]

Arjun Teotia[‡]

May 25, 2022

Abstract

Studying the relationship between smoking and obesity, the leading causes of preventable deaths in the U.S., helps assess the potential unintended consequences of policy efforts to reduce smoking. Because existing literature is mixed among studies using experimental and observational data, we investigate the role of misreporting in observational data. We use the Behavioral Risk Factor Surveillance System, cigarette taxes to instrument for smoking, and survey completion to instrument for misreporting. Starting with the seminal two-stage least squares (2SLS) approach, our estimates similarly suggest quitting smoking substantially reduces BMI. However, the relationship between cigarette taxes and BMI has shrunk over time, and the 2SLS estimates are sensitive to specification, functional form, and misreporting. Accounting for misreporting using the [Nguimkeu et al. \(2019\)](#) 2-step estimator yields estimates consistent with the experimental literature; quitting smoking modestly increases BMI. Our preferred estimate suggests reduced smoking accounts for 6% of the concurrent rise in obesity.

*IZA, NBER, Georgia State University

[†]Georgia State University, corresponding author (kteltser@gsu.edu)

[‡]Georgia State University

1 Introduction

Smoking and obesity are the two leading causes of preventable deaths in the United States.¹ Almost 40% of adults in the United States are obese (Hales et al. (2017)), while roughly 14% of adults in the U.S. smoke according to the CDC.² In Figure 1, we show that smoking decreased from roughly 22% to 15% between 2002 and 2017 among Behavioral Risk Factor Surveillance System (BRFSS) respondents, while average body mass index (BMI) rose from roughly 26.75 to 28.25 during the same time period. While these trends do not present causal evidence, they raise the question of whether the decline in smoking has contributed to rising obesity. Because smoking is subject to heavy government intervention, accurately estimating the effect of smoking on obesity is important in determining the extent of unintended costs or benefits of such intervention.

Using Grossman’s human capital model of health, we can consider health a capital stock that depreciates over time but can be replenished by investment (Grossman, 1972). Assuming an individual starts at her optimal body weight, both smoking and weight gain represent disinvestments. Cigarettes contain nicotine, which suppresses appetite and increases one’s metabolic rate. Thus, smoking cigarettes tends to reduce body weight, all else equal. However, smoking also reduces lung capacity, which may restrict physical activity and thereby increase body weight. Because the causal effect of smoking on obesity is theoretically ambiguous, determining the true relationship is an empirical question.

There exists a large literature examining the causal relationship between smoking and obesity, though the findings are quite mixed. Using BRFSS data from 1984 to 1999, Chou et al. (2004) suggest that reduced smoking increases body mass index (BMI) by finding increasing cigarette prices have a large positive effect on BMI. Using BRFSS data from 1984 to 2002/2005, Gruber and Frakes (2006) and Courtemanche (2009) examine this relationship more closely using cigarette taxes instead of cigarette prices.³ The authors find that an increase in cigarette taxes (i.e. reduced

¹See https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm and <https://www.commonwealthfund.org/blog/2018/rising-obesity-united-states-public-health-crisis>

²See https://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/index.htm, last accessed February 19, 2020

³Note that Courtemanche (2009) actually uses BRFSS data to examine the robustness of his central estimates obtained using the National

smoking) *reduces* BMI. However, several other studies around the same time use different observational data or approaches to causal identification to suggest that reduced smoking increases BMI (Baum, 2009; Fang et al., 2009; Liu et al., 2010). Moreover, Courtemanche et al. (2018) also find that reduced smoking increases BMI using a randomized trial of smoking cessation treatments and clinically-measured carbon monoxide levels.

The mixed findings highlight the complexity of identifying the causal effect of smoking on BMI. In particular, two concerns stand out from the literature. First, the estimates are sensitive to different approaches to identification and choice of specification. Second, these estimates are sensitive to the source of data, which raises the concern that results obtained from self-reported data (as opposed to clinically-measured experimental data) are biased and/or inefficient due to systematic misreporting. For example, studies have found that accounting for misreporting can significantly impact estimates of marijuana usage, as well as the magnitude and direction of estimated effects of nutritional assistance programs on BMI and obesity (Greene et al., 2017; Almada et al., 2016; Nguimkeu et al., 2019). A large body of medical research has revealed widespread underreporting based on comparisons of self-reported smoking to biochemical indicators of tobacco intake (e.g., Gorber et al., 2009; Nesson, 2017). To our knowledge, no one has previously examined how misreporting impacts estimates of the causal effect of smoking on BMI.

In this paper, we aim to revisit the mixed findings in the existing literature and provide key insights for future researchers assessing the relationship between smoking and obesity. To do this, we use BRFSS data from 2002 to 2017 to examine how specification choice and accounting for misreporting affect estimates of the effect of smoking on BMI. In particular, we employ a variety of linear and nonlinear instrumental variables approaches, using cigarette taxes as plausibly exogenous shifters of smoking behavior.

First, we replicate the findings of Gruber and Frakes (2006) and Courtemanche (2009) using our more recent data, obtaining two-stage least squares (2SLS) estimates suggesting that quitting smoking reduces BMI by 4.2 (27 pounds for someone who is 5'7" tall and 180 pounds). That said,

Longitudinal Survey of Youth. Indeed, the author finds that both data sources yield similar estimates.

this 2SLS estimate is quite imprecise, and any conventional confidence interval would fail to rule out the possibility that quitting smoking increases BMI. Moreover, we demonstrate that the 2SLS estimates and standard errors are quite sensitive to the inclusion/exclusion of covariates and fixed effects.

Second, we introduce the 2-Step estimation procedure proposed by [Nguimkeu et al. \(2019\)](#), which uses a nonlinear first stage to predict smoking status and accounts for misreporting. We obtain estimates consistent with the experimental literature in both sign and magnitude. That is, we find quitting smoking *increases* BMI by 1.06, or about 7 pounds for someone who is 5’7” tall and 180 pounds. To show the intermediate impact of using a nonlinear first stage to predict smoking status, we present estimates from both a control function and nonlinear-fits-as-instruments approach. Here, we find estimates of roughly the same or larger magnitude as 2SLS but opposite-signed, suggesting again that quitting smoking increases BMI.

One possible explanation for the discrepancy between our 2SLS estimates and those from earlier studies is that the causal effect of smoking on obesity has changed over time. However, the more likely explanation is that 2SLS estimates are more sensitive than previously thought. Beyond finding large sensitivity to the inclusion/exclusion of covariates and fixed effects, we show in supplemental analyses that our 2SLS estimate is extremely sensitive to adding our measure of misreporting (see Section 3 for more detail) as a second instrument (going from 4.2 to -20.1). This is consistent with [Almada et al. \(2016\)](#), who find that 2SLS estimates of the effect of SNAP on obesity fall well outside of nonparametric (“Manski”) bounds in the presence of misreporting. Meanwhile, our control function and nonlinear-fits-as-instruments specifications yield essentially identical estimates with or without the inclusion of this second instrument.

Overall, we demonstrate that accounting for misreporting and leveraging nonlinearity in identification affects the causal estimates of smoking on BMI by improving precision, flipping the sign of the relationship, and bringing the magnitudes much closer to medical estimates. From 2002-17, our main estimate (i.e. quitting smoking increases BMI by 1.06) suggests that the decline in smoking explains roughly 6% of the concurrent rise in obesity, suggesting that increased obesity may

be an unintended consequence of anti-smoking policy efforts. Moreover, our 2-Step heterogeneity estimates suggest that quitting smoking increases BMI more for those aged 18-40 than 41-65, more for men than women, and more for those in the northeast and midwest regions of the U.S. relative to those in the southern and western regions.

One limitation of our work is that it cannot speak to any discrepancies in the existing literature regarding the reduced-form impact of cigarette prices or taxes on obesity, since misreported smoking status has little to no bearing on changes in reported BMI. However, we do obtain substantially smaller estimates of the reduced-form effect of cigarette taxes on BMI from 2002 to 2017, by roughly an order of magnitude compared to [Baum \(2009\)](#), [Chou et al. \(2004\)](#), and [Courtemanche \(2009\)](#), suggesting that this relationship has shrunk over time.

Ultimately, our work highlights that additional research on the causal relationship between smoking and obesity is warranted, and provides researchers with some key lessons to consider in future work. The rest of the paper is organized as follows. Section 2 provides a more detailed review of the literature. Section 3 describes our data. Section 4 develops our various approaches to estimation and presents our results. Section 5 concludes.

2 Literature Review

Instrumental variables (IV) approaches are common in the literature on the relationship between smoking and body weight. This is because smoking behavior may be endogenous, as individuals generally self-select into smoking. There are also concerns about reverse causality. That is, body weight may affect an individual's decision of whether or not to smoke ([Cawley et al., 2004](#); [Rees and Sabia, 2010](#)). The literature has primarily used cigarette prices and taxes as plausibly exogenous instruments, as they shift the cost of smoking but do not seem to have a direct causal relationship with BMI.

In Table 1, we catalog the main estimates from earlier literature on the relationship between smoking and BMI. First, [Chou et al. \(2004\)](#) use BRFSS data from 1984 to 1999 to estimate the

determinants of BMI and obesity, and find large positive effects of cigarette prices. Assuming cigarette prices affect individuals only through smoking, these results suggest that reductions in smoking contribute to increases in body weight. In contrast, using BRFSS data from 1984 to 2002 and cigarette taxes instead of prices, [Gruber and Frakes \(2006\)](#) examine this question more closely and find that reduced smoking reduces BMI and obesity rates. The authors propose two possible explanations for why their results differ from those of [Chou et al. \(2004\)](#). First, taxes may be better instruments than prices, as variation in prices may be driven by factors affecting both BMI and demand for cigarettes. Second, [Gruber and Frakes \(2006\)](#) use month-year fixed effects rather than quadratic time trends to more flexibly control for unobservable location-invariant shocks over time. However, the authors caution that their IV estimates may be implausibly large, as they suggest that individuals who quit smoking are 56% less likely to be obese. [Courtemanche \(2009\)](#) revisits [Gruber and Frakes \(2006\)](#) using both BRFSS and the National Longitudinal Survey of Youth (NLSY), and also finds that reduced smoking reduces BMI and obesity rates. Moreover, the author shows that this result is robust to a range of alternative linear specifications. [Wehby and Courtemanche \(2012\)](#) obtain similar results using BRFSS and a quantile regression approach in their examination of heterogeneity in the effects of cigarette prices on BMI by race, education, age, and sex.

Other studies around the same time, however, challenge these findings. [Baum \(2009\)](#) use the NLSY for years 1979 through 2002 to estimate the body-weight effects of real cigarette prices and taxes. Using a standard differences-in-differences approach, they find that an increase in the cost of cigarettes (i.e. a reduction in smoking) increases BMI and obesity rates. [Fang et al. \(2009\)](#) use data from the China Health and Nutrition Survey (CHNS), leveraging number of cigarettes smoked and the local per-pack price of the most popular cigarette brand as IVs. Similarly, they find that reduced smoking increases BMI, though they find small and statistically insignificant effects on obesity rates. [Liu et al. \(2010\)](#) use worksite smoking bans along with BRFSS data from 1998 to 2006 to find further evidence suggesting reduced smoking increases obesity rates. In more recent work, [Courtemanche et al. \(2016\)](#) use BRFSS data from 1990 to 2010 to examine the

effects of a wide range of economic factors on obesity, finding a negative effect of prices on BMI when only regressing BMI on cigarette prices, but the coefficient turns positive when including the full set of economic factors as covariates. [Pieroni and Salmasi \(2016\)](#) use the British Household Panel Survey from 2004 and 2006 to estimate that quitting smoking increases BMI by a relatively modest amount. [Entezarkheir et al. \(2016\)](#) use the Canadian Community Health Surveys from 2003 and 2005, and find that cigarette tax increases are associated with an increase in the rate of obesity. [Amialchuk et al. \(2018\)](#) use the 1996 to 2008 waves of the Belarusian Household Survey of Income and Expenditure to estimate a relatively large negative relationship between smoking and BMI.

Moreover, a more recent study by [Courtemanche et al. \(2018\)](#) uses clinical measurements from an experimental setting to provide additional consistent evidence. In particular, they use data from a randomized trial of smoking cessation treatments. Rather than relying on self-reported smoking, they observe clinically-measured carbon monoxide levels. They find that, in the short run, quitting smoking leads to a gain of 10-11 pounds at average height. In the long run, this effect increases to 11-12 pounds. We view these as the most credible estimates to date. The main limitation of [Courtemanche et al. \(2018\)](#) is that the study focuses on individuals who were trying to quit smoking in 1990-1994, which may cast doubt on whether similar estimates would be obtained in more recent years.

While not an exhaustive compilation, the mixed evidence from the body of literature discussed above highlights the challenges inherent in estimating the causal relationship between smoking and body weight. Misreporting is one significant challenge that the literature on smoking and obesity has seemingly not yet tackled. [Nesson \(2017\)](#) finds about 46% of adolescent smokers misreport smoking status. Subjects in the study knew ahead of time that they would later be clinically tested for evidence of smoking activity, suggesting 46% may be an underestimate of misreporting in regular survey data like the BRFSS. That said, adolescents may also be more likely to misreport than adults. In any case, this level of misreporting can cause a large bias in the second stage. [Nguimkeu et al. \(2019\)](#) and [Brachet \(2008\)](#) both provide insight on this issue, in that the authors discuss how

expansion bias occurs when misreporting is fixed and exogenous. Unfortunately, it is not straightforward to quantify how large this bias will be given a particular level of misreporting. [Almada et al. \(2016\)](#) also speak to this issue, examining the relationship between supplemental nutrition assistance (SNAP) and obesity in the presence of misreporting. The authors find that traditional linear IV estimates fall outside of the Manski bounds, providing further evidence that traditional linear IV can be quite unreliable in the presence of misreporting. More generally, [Mogstad and Wiswall \(2010\)](#) show the fragility of traditional linear IV approaches when the underlying relationship is nonlinear, while [Guo and Small \(2016\)](#) show that a control function IV approach can be much more efficient than two-stage least squares under certain conditions.

In the following sections, we revisit the mixed literature with the aim of demonstrating that specification choice, functional form, and misreporting can have dramatic impacts on the estimated causal relationship between smoking and obesity.

3 Data

We start with repeated cross-sectional data from the 2002 through 2017 waves of the BRFSS, which is a health-focused telephone survey conducted by the CDC. The survey collects information on health-related risky behaviors among U.S. residents, as well as chronic health conditions and use of preventative services. The BRFSS was first established in 1984, with 12,258 participants across 14 states. In 2017, the data include approximately 430,000 participants from all 50 states and the District of Columbia. Starting in 2002, the BRFSS contains information on interviewer performance and whether a respondent fully or partially completed the interview. Because we use this information to account for possible misreporting, we restrict our main estimation sample to the years 2002-2017.

In particular, we use the core module that includes self-reported height, weight, demographics, tobacco use, health status, and chronic health conditions. Following [Gruber and Frakes \(2006\)](#), we restrict our sample to respondents under 65 years of age. After also dropping observations with

missing data for our key variables of interest, our sample contains roughly 3.6 million observations. We calculate BMI using respondents' self-reported height and weight, removing extreme outliers by dropping the bottom and top 1% of BMI values, which leaves us with a range of 17.75 to 46.39. Using CDC guidelines, we define a current smoker as one who reported smoking at least 100 cigarettes during their lifetime, and, at the time of participation, reported that they currently smoke either some days or every day. We also use several additional individual-level covariates, including marital status, race, education, and income. Similar to [Gruber and Frakes \(2006\)](#), we include age-by-gender interaction terms.

We supplement our BRFSS data with information on state excise cigarette taxes from the CDC's Tax Burden on Tobacco (TBOT) dataset. The TBOT reports tax rates at the end of October every year, along with the exact date(s) of any rate changes, which we use to determine monthly tax rates. We convert nominal tax rates to 2017 dollars using the Consumer Price Index for All Urban Consumers (CPI-U). Because the BRFSS asks individuals about their behavior over the past 30 days, we follow the previous literature in assigning tax rates to individuals based on the month preceding their survey interview. Finally, we include a time-varying measure of state-level unemployment from the Bureau of Labor Statistics (BLS) to account for economic conditions that may be correlated with body weight, smoking, and/or cigarette taxes.

In [Table 2](#), we report the summary statistics for our main estimation sample derived from the 2002-2017 BRFSS files, which keeps only the respondents for whom all of our control and outcome variables of interest are non-missing. In this sample, 21% of respondents are identified as smokers. The average BMI among smokers is 27.18, which is lower than the average non-smoker BMI of 27.77. In addition, we see that 78% of individuals are white, 8% are black, and 7% are hispanic. In addition, 43% of individuals are male, 27-28% have a high school and/or technical degree, and 38% are college graduates. Note that, in our analyses, we weight by the stratified sampling weights provided by the CDC. Using survey weights is important in this context, as BRFSS oversamples different population groups across different years depending on the guidance from the CDC.

In the sample, the mean (real) cigarette excise tax is \$2.22. The highest observed nominal tax is \$4.35 in New York and Connecticut in 2017, while the lowest in 2017 is \$0.17 in Missouri (see [Table 3](#)). In addition to substantial across-state variation in tax magnitudes, taxes vary substantially across time within states. Between 2002 and 2017, there were 132 nominal tax increases.

Finally, the variable we use to account for misreporting is called “Final Disposition” in the BRFSS data. This indicates whether the respondent fully completed their interview, and we use it to predict whether an individual is less likely to answer questions truthfully. We find that approximately 7% of respondents in our sample only partially complete their interview.

4 Estimation and Results

In this section, we start by obtaining naïve OLS estimates of the relationship between smoking and BMI via OLS, and then estimate a variety of IV specifications. Throughout, we follow the previous literature in specifying the following structural linear model of the relationship between BMI and smoking:

$$\text{BMI}_{ist} = \alpha S_{ist} + X_{ist}\beta + \tau_t + \mu_s + \epsilon_{ist} \quad (1)$$

where S indicates current smoking status for individual i in state s and month-year t . X is a vector of individual- and state-level covariates, τ_t are time fixed effects,⁴ μ_s are state fixed effects, and ϵ_{ist} is an idiosyncratic error term. X includes individuals’ income, race, marital status, education, age, gender, an age-by-gender interaction term, and state unemployment rate from the BLS.

In the first row, second panel of [Table 4](#), we find that the naïve OLS estimate of the relationship between smoking and BMI is negative. For OLS, and all our other estimation approaches, we present three specifications. Column 1 presents the results when we only include state and quarter-year fixed effects. Column 2 presents the results when we only include the set of covariates (income, race, marital status, educational attainment, age-by-gender dummies, and state unemployment rate). Column 3 presents the results when including both the fixed effects and co-

⁴Note that we use quarter-year, not month-year, for our time fixed effects throughout to improve computational feasibility for the nonlinear portions of our analyses.

variates. In our preferred specification, column 3, we estimate that quitting smoking is associated with a BMI increase of 0.7 (2.5% relative to average BMI of 27.65). For someone who is 5' 7" tall and 180 pounds (BMI = 28.2), this corresponds to an increase of roughly 4.5 pounds.

Our first IV approach is a standard two-stage least squares (2SLS) approach, which follows [Gruber and Frakes \(2006\)](#) closely. Next, we implement a 2-step estimator proposed by [Nguimkeu et al. \(2019\)](#) to account for the possibility of misreporting, which entails using a nonlinear first stage. Finally, to show the reader the impact of using a nonlinear first stage without accounting for misreporting, we obtain estimates from two alternate IV approaches. The first is a corrected version of the “forbidden regression.” The second is a control function approach. Following the large economic literature on smoking, we use cigarette taxes to instrument for smoking status throughout these analyses. In addition, we weight our regressions throughout by the BRFSS-provided stratified sampling weights.⁵

4.1 Two-Stage Least Squares

In this approach, following [Gruber and Frakes \(2006\)](#), the first stage is specified as a linear probability model:

$$S_{ist} = \psi \text{Tax}_{st} + X_{ist}\phi + \tau_t + \mu_s + \eta_{ist} \quad (2)$$

Tax is the identifying instrument, which reflects the real excise cigarette tax (in 2017 \$). The corresponding second stage is given by:

$$\text{BMI}_{ist} = \alpha \hat{S}_{ist} + X_{ist}\beta + \tau_t + \mu_s + \epsilon_{ist} \quad (3)$$

where \hat{S} is predicted smoking status from the first stage, and α is the causal parameter of interest.

[Table 4](#) presents the results from the analysis using 2SLS. In the first stage (row 1, top panel, column 3), we estimate that a \$1 increase in the real cigarette excise tax significantly reduces the probability of being a smoker by 0.6 percentage points. We also present the F-statistic, which is

⁵See appendix for non-weighted estimates.

16.4 for this specification, suggesting that cigarette taxes are indeed a strong predictor of smoking behavior. Note that we do not find a statistically significant direct relationship between BMI and cigarette taxes, where the reduced form estimate presented in column 3 of the middle panel suggests that a \$1 tax increase statistically insignificantly reduces BMI by 0.025. In row 2, column 3 of the bottom panel, our 2SLS estimate suggests that quitting smoking reduces BMI by roughly 4.2 (or a reduction of 27 pounds for someone 5' 7" tall and 180 pounds). While the standard error of the 2SLS estimate is roughly 50% larger than the coefficient, the magnitude is in line with the estimate of 5.70 (s.e. = 2.91) from [Gruber and Frakes \(2006\)](#) using BRFSS data from 1984-2002 and our replication of that estimate of 5.34 (s.e. = 3.25). That said, as pointed out by the authors themselves, it still seems to be unreasonably large when compared to the rest of the literature.

In addition, it is also worth noting that the 2SLS estimates are relatively unstable across specifications. When including fixed effects but no covariates, the 2SLS estimate is similar but smaller (3.49). However, when including covariates and excluding fixed effects, the sign flips and the magnitude is much smaller (-0.79, more in line with the stable “naïve” OLS estimates of -0.4 to -0.7). Moreover, the 2SLS standard errors reveal that the point estimates are quite imprecise, encompassing a wide range of both positive and negative values within the 95% confidence interval. For example, the main estimate of 4.2 has a standard error of roughly 6.3. Going two standard deviations in both directions yields a 95% confidence interval of [-8.404, 16.804]. In the results presented in the following subsections, we see that the preferred 2-step, NFI, and CF estimates all fall within this interval. This imprecision appears to be driven, not by a weak first stage, but by a weak reduced form relationship between cigarette taxes and BMI. Estimating our preferred specification suggests that a \$1 increase in cigarette taxes reduces average BMI by 0.025 with a 95% confidence interval of [-0.109, 0.059].

In the following subsections, we show (1) the sign of the estimates is consistently negative when using a nonlinear first stage, (2) conditional on including covariates, the results are not nearly as sensitive to the inclusion of fixed effects when using a nonlinear first stage, and (3) using a nonlinear first stage with the full set of controls yields much more precise IV estimates, roughly an

order of magnitude smaller than the 2SLS counterpart. Moreover, we show that the results are stable across all three specifications when using a nonlinear first stage that accounts for misreporting, ranging only from approximately -0.50 to -1.06. It is possible that the gain in precision is a byproduct of specifying a more accurate first stage functional form, which perhaps improves the precision of the resulting second stage IV estimator.

4.2 2-step estimator

Next, using a 2-step estimator allows us to account for misreporting of smoking in the BRFSS data, following [Nguimkeu et al. \(2019\)](#). Suppose the outcome variable of interest, BMI, is a function of correctly-measured exogeneous covariates X and the (true) smoking indicator, S^* . In particular, we specify:

$$BMI_{ist} = \alpha S_{ist}^* + X_{ist}\beta + \tau_t + \mu_s + \epsilon_{ist}. \quad (4)$$

where α is the key parameter of interest.

We model true smoking status (S^*) as a function of X and the exogenous instrument Tax :

$$S_{ist}^* = \mathbf{1}(\theta Tax_{st} + X_{ist}\beta + \tau_t + \mu_s + v_{ist} \geq 0). \quad (5)$$

However, we only observe self-reported smoking status, S , which is a surrogate of one's true smoking status. Thus, we can consider $S_i = S_i^* d_i$, where d is an indicator of misreporting. If an individual truthfully reports their smoking status, $d = 1$ (0 otherwise). Note that $S = S^*$ for true non-smokers regardless of misreporting. That is, we assume a true non-smoker would not falsely report smoking. Next, we model d as a function of observable covariates X and disposition $Disp$:

$$d_{ist} = \mathbf{1}(\rho Disp_{ist} + X_{ist}\gamma + \tau_t + \mu_s + u_{ist} \geq 0) \quad (6)$$

such that:

$$S_{ist} = S_{ist}^* d_{ist} \\ = \mathbf{1}(\theta \text{Tax}_{st} + X_{ist}\beta + \tau_t + \mu_s + v_{ist} \geq 0 ; \rho \text{Disp}_{ist} + X_{ist}\gamma + \tau_t + \mu_s + u_{ist} \geq 0). \quad (7)$$

In the first step, we estimate equation 7 using a partial observability probit model, which allows us to predict each individual's true smoking status, \hat{S}^* . In the second stage, we estimate equation 4, substituting \hat{S}^* in for S^* . The resulting α is the 2-step estimate of the true causal effect of smoking on obesity when accounting for endogenous misreporting.

The results of this approach, presented in row 2 of the top panel of Table 4, yield very similar first-stage estimates of the effect of cigarette taxes on smoking. In our preferred specification (column 3), the first-stage F-statistic is 36.95, again suggesting that cigarette taxes are strong predictors of smoking. Turning to row 3 of the bottom panel, we find a negative relationship between smoking and obesity. However, here we find that quitting smoking leads to an increase in BMI of only 1.06 (about 7 pounds for someone 5' 7" tall and 180 pounds). This estimate is much closer to the experimental 2SLS estimate obtained by Courtemanche et al. (2018), who find that quitting smoking increases BMI by 2.2 (about 14 pounds) when using clinically-measured carbon monoxide levels to determine smoking behavior among participants in a smoking cessation randomized trial. Moreover, we find that this 2-step approach yields a much more stable pattern of estimates across our three specifications, where the estimates only range between -0.50 and -1.06. Our preferred estimate of -1.06 is also considerably more precise than its 2SLS counterpart, with a 95% confidence interval of only [-2.057, -0.061].

4.3 Nonlinear First Stage Without Accounting for Misreporting

In the following two sub-subsections, we present to the reader the intermediate impact of using a nonlinear first stage without accounting for misreporting. In particular, we obtain estimates from two alternate IV strategies: nonlinear-fits-as-instruments and a control function approach.

4.3.1 Nonlinear-Fits-as-Instruments

First, we would like to estimate the relationship between smoking and BMI by using a probit model for the first stage to obtain predicted smoking status, and then substituting predicted smoking status for observed smoking status in a linear second stage. However, this was famously dubbed the “forbidden regression” by Hausman, since only a linear first stage estimated via OLS is “guaranteed to produce first-stage residuals that are uncorrelated with fitted values and covariates” (Angrist and Pischke, 2008).

To correct for this, we use the nonlinear-fits-as-instruments (hereafter, NFI) approach suggested by Newey (1990). The NFI approach has three stages, which we describe below as stages zero, one, and two. Stage zero is given by:

$$S_{ist} = \psi' \alpha x_{st} + X_{ist} \phi + \tau_t + \mu_s + \eta_{ist}, \quad (8)$$

which we estimate as a probit model to obtain predicted smoking status \hat{S} . Then, in stage one, we estimate S as a function of \hat{S} along with our set of controls:

$$S_{ist} = \theta \hat{S}_{ist} + X_{ist} \gamma + \tau_t + \mu_s + u_{ist}. \quad (9)$$

Finally, we estimate stage two of the NFI model, where BMI is specified as a function of predicted smoking from stage one, \tilde{S} , and the set of controls:

$$BMI_{ist} = \alpha \tilde{S}_{ist} + X_{ist} \beta + \tau_t + \mu_s + \epsilon_{ist}. \quad (10)$$

Note that this approach identifies α using both taxes and nonlinearities in stage zero. While using nonlinearities as identifying information may be somewhat undesirable, Newey (1990) shows that the resulting estimates are more efficient when the probit model is a better approximation of the first-stage conditional expectation function.

Turning to Table 4, row 3, column 3 of the top panel reveals a very similar first stage relation-

ship as in 2SLS. Specifically, the estimate suggests that a \$1 increase in the real cigarette excise tax reduces the probability of being a smoker by 0.6 percentage points. The first stage F-statistic is 11.38, again, suggesting that cigarette taxes are a strong predictor of smoking. However, turning to row 4, column 3 of the bottom panel, we find that the estimated effect is similar in magnitude to 2SLS but the direction of the relationship is flipped. That is, our NFI estimate suggests that quitting smoking increases BMI by 6.2 (roughly 40 pounds for someone 5' 7" tall and 180 pounds). In this specification, as opposed to 2SLS, the standard errors are much smaller and the coefficient is highly statistically significant. That said, the magnitude still appears to be unreasonably large relative to the rest of the literature. While the corresponding NFI estimate when we include fixed effects and exclude covariates is particularly unreasonably large (-20.9) and imprecise (s.e. = 16.2), when we include covariates the estimates are not as sensitive as 2SLS to the inclusion of fixed effects (-4.5 without fixed effects vs. -6.2 with fixed effects).

4.3.2 Control Function

Second, we estimate the causal relationship of interest using a nonlinear (probit) first stage via the control function (hereafter, CF) approach. Again, the first stage is given by:

$$S_{ist} = \psi T \alpha x_{st} + X_{ist} \phi + \tau_t + \mu_s + \eta_{ist}. \quad (11)$$

After estimating this probit model, we obtain the residuals (Res) and use them as a covariate in the second stage. The second stage is given by:

$$BMI_{ist} = \alpha S_{ist} + \delta Res_{ist} + X_{ist} \beta + \tau_t + \mu_s + \epsilon_{ist} \quad (12)$$

where S is self-reported smoking status, Res are the residuals from the first stage, along with the usual set of additional controls. The estimated α from the second stage is our causal parameter of interest.

The results presented in row 3, column 3 of the top panel of [Table 4](#) reflect the first stage esti-

mates for both the NFI and CF approaches, as they both use the exact same first stage specification. Turning to row 5, column 3 of the bottom panel, we see that the estimate is quite similar to the NFI estimate (negative relationship) but slightly larger in magnitude (-7.4 vs. the NFI estimate of -6.2). As in the NFI specification, the standard error is much smaller than that of the 2SLS estimate and the coefficient of interest is highly statistically significant. Again, while the corresponding CF estimate when we include fixed effects and exclude covariates is unreasonably large (-11.3) and imprecise (s.e. = 8.6), when we include covariates the estimates are not as sensitive as 2SLS to the inclusion of fixed effects (-5.5 without fixed effects vs. -7.4 with fixed effects).

Even though the sign flips when using a nonlinear first stage, which is in line with the experimental literature, the estimates are still unreasonably large compared to the rest of the literature on the relationship between smoking and BMI. This highlights the potential importance of accounting for misreporting in smoking status via the 2-step procedure, which has a substantial impact on the stability of the estimates while bringing them back within a reasonable range of values.

4.4 Including Disposition as a Second Instrument in 2SLS, NFI, and CF

In this subsection, we re-estimate our preferred 2SLS, NFI, and CF approaches including both cigarette taxes and disposition (i.e. our proxy for misreporting) as instrumental variables. This helps us address the concern of whether our 2-Step estimates are simply the byproduct of introducing a second instrument rather than addressing misreporting specifically. We present these results in [Table 5](#). The notable results are the following. First, including disposition as a second instrument has virtually no effect on the NFI and CF estimates. The NFI estimate is -6.481 and the CF estimate is -7.694, compared to -6.210 and -7.366, respectively, in [Table 4](#). The standard errors are also very similar. However, including disposition as a second instrument has an enormous impact on the 2SLS estimate, yielding a much larger and opposite-signed coefficient of -20.131. Despite the large standard error of 9.643, this estimate is statistically significant at the 5% level. This result further highlights the extreme sensitivity, and thus relative unreliability, of the fully-linear IV approach in examining the effect of smoking on BMI using survey data and cigarette taxes as an

instrument.

4.5 Heterogeneity

Finally, we use the 2-step estimator to examine heterogeneity in the effects of smoking on BMI across age, gender, and U.S. census region. While we will focus primarily on presenting the 2-step results, note that we also present the results of OLS, 2SLS, NFI, and CF estimation for comparison purposes.

Turning first to Table 6, our 2-step estimates suggest that cigarette taxes have a slightly larger impact on smoking among respondents aged 41-65: a reduction of 1.1 percentage points in the probability of being a smoker for a \$1 increase in taxes compared to 0.8 points for those aged 18-40. In terms of the effects of smoking status on BMI, however, we find that quitting smoking has a larger positive impact on BMI among the 18-40 age group: an increase of 4.2 (27 pounds for someone 5' 7" and 180 pounds) compared to only 0.8 (5 pounds) among those aged 41-65.

Next, looking at Table 7, our 2-step estimates suggest that cigarette taxes have a substantially larger impact on smoking among female respondents: a 1.2 percentage point smoking reduction compared to 0.6 for men. Interestingly, we find that quitting smoking results in a significantly larger weight gain for men than women: 6.9 BMI points for men compared to a statistically insignificant decrease of 1.3 BMI points for women.

Finally, in Table 8, we examine regional differences in the relationship between smoking and BMI. Looking in row 3 of the bottom panel, we find that quitting smoking increases BMI by 3.2 to 4.5 in the northeast and midwest, while we find no statistically discernable relationship among respondents in the southern and western regions of the United States.

5 Conclusion

Because smoking and obesity are the two leading causes of preventable deaths in the United States, and are thus subject to heavy government intervention, understanding the effect of smoking on obe-

sity is important in determining the costs or benefits of such intervention. To date, the literature examining this relationship has been quite mixed, with some studies finding a positive effect and others finding a negative effect. In this paper, we show that changing the functional form, accounting for misreporting, and including/excluding covariates and fixed effects can dramatically affect estimates of the causal relationship between smoking and obesity.

In particular, we show that using a nonlinear first stage to estimate the relationship between cigarette taxes and smoking status yields a negative relationship between smoking and BMI, which is consistent with the experimental literature and some of the earlier literature using observational data. Moreover, we show that accounting for misreporting and using a nonlinear first stage with observational data yields estimates of similar sign and magnitude to this body of work. For example, our preferred estimate suggests that quitting smoking increases BMI by 1.06 (about 7 pounds for someone 5' 7" and 180 pounds), which is quite similar to the experimental estimate of 2.2 (14 pounds) from [Courtemanche et al. \(2018\)](#), but quite different from the estimated *decrease* in BMI of 4.2 to 5.7 (27 to 36 pounds) when following the approach of [Gruber and Frakes \(2006\)](#) using 2SLS and observational data.

Using our estimation sample and incorporating survey weights, the proportion of smokers declined from 25.4% in 2002 to 18.8% in 2017 while BMI rose from 26.7 to 27.9. Multiplying the 6.6 percentage point decline in smoking by our preferred estimate of 1.06 yields a BMI increase of 0.07, which explains roughly 6% of the overall BMI increase of 1.2 across the same time period. This suggests increased obesity may be an unintended consequence of anti-smoking policy efforts.

While our work cannot speak to any discrepancies in the existing literature regarding the reduced-form impact of cigarette prices or taxes on obesity, we do obtain substantially smaller estimates of the reduced-form effect of cigarette taxes on BMI from 2002 to 2017 suggesting that the relationship has shrunk over time. Note, however, that a negative relationship between cigarette prices/taxes and BMI does not necessarily imply a positive effect of smoking on weight. In particular, cigarette taxes may affect BMI through channels other than smoking (i.e. using cigarette taxes as an instrument for smoking may not satisfy the exclusion restriction). Because smoking

is relatively inelastic due to the addictive nature of cigarettes, a large tax increase likely generates a very modest decrease in smoking. Thus, only a small share of smokers is subject to any corresponding BMI gain. The remaining smokers simply experience a decrease in real income, which can have a causal effect on weight (for example, see [Cawley \(2015\)](#) for a review of the literature on the economic causes of obesity).

Ultimately, our work shows how certain functional form choices and failing to account for misreporting might result in biased and/or inefficient estimates when using survey data and instrumental variables methods. These considerations have received little attention in the literature on the effects of smoking. As such, a key contribution of this paper is to demonstrate how researchers might effectively address functional form and misreporting issues in future work.

Bibliography

- Almada, L., McCarthy, I., and Tchernis, R. (2016). What can we learn about the effects of food stamps on obesity in the presence of misreporting? *American Journal of Agricultural Economics*, 98(4):997–1017.
- Amialchuk, A., Bornukova, K., and Ali, M. M. (2018). Will a decline in smoking increase body weights? evidence from belarus. *Eastern Economic Journal*, 44(2):190–210.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Baum, C. L. (2009). The effects of cigarette costs on bmi and obesity. *Health Economics*, 18(1):3–19.
- Brachet, T. (2008). Maternal smoking, misclassification, and infant health. *SSRN Working Paper*.
- Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43:244–268.
- Cawley, J., Markowitz, S., and Tauras, J. (2004). Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation. *Journal of health economics*, 23(2):293–311.
- Chou, S.-Y., Grossman, M., and Saffer, H. (2004). An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. *Journal of health economics*, 23(3):565–587.
- Courtemanche, C. (2009). Rising cigarette prices and rising obesity: Coincidence or unintended consequence? *Journal of Health Economics*, 28(4):781–798.
- Courtemanche, C., Tchernis, R., and Ukert, B. (2018). The effect of smoking on obesity: Evidence from a randomized trial. *Journal of health economics*, 57:31–44.
- Courtemanche, C. J., Pinkston, J. C., Ruhm, C. J., and Wehby, G. L. (2016). Can changing economic factors explain the rise in obesity? *Southern Economic Journal*, 82(4):1266–1310.
- Entezarkheir, M., Sen, A., and Wilson, A. (2016). Obesity, smoking, and cigarette taxes: Evidence from the canadian community health surveys. *SSRN Working Paper No. 2808127*.
- Fang, H., Ali, M. M., and Rizzo, J. A. (2009). Does smoking affect body weight and obesity in china? *Economics & Human Biology*, 7(3):334–350.
- Gorber, S. C., Schofield-Hurwitz, S., Hardt, J., Levasseur, G., and Tremblay, M. (2009). The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine & Tobacco Research*, 11(1):12–24.
- Greene, W., Harris, M. N., Srivastava, P., and Zhao, X. (2017). Misreporting and econometric modelling of zeros in survey data on social bads: An application to cannabis consumption. *Health economics*.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255.
- Gruber, J. and Frakes, M. (2006). Does falling smoking lead to rising obesity? *Journal of health economics*, 25(2):183–197.
- Guo, Z. and Small, D. S. (2016). Control function instrumental variable estimation of nonlinear causal effect models. *The Journal of Machine Learning Research*, 17(1):3448–3482.
- Hales, C., Carroll, M., Fryar, C., and Ogden, C. (2017). Prevalence of obesity among adults and youth: United states, 2015-2016. *NCHS data brief*, (288):1–8.

- Liu, F., Zhang, N., Cheng, K.-W., and Wang, H. (2010). Reduced smoking and rising obesity: Does smoking ban in the workplace matter? *Economics Letters*, 108(3):249–252.
- Mogstad, M. and Wiswall, M. (2010). Linearity in instrumental variables estimation: problems and solutions. *IZA Discussion Paper No. 5216*.
- Nesson, E. (2017). The impact of tobacco control policies on adolescent smoking: comparing self-reports and biomarkers. *American Journal of Health Economics*, 3(4):507–527.
- Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, pages 809–837.
- Nguimkeu, P., Denteh, A., and Tchernis, R. (2019). On the estimation of treatment effects with endogenous misreporting. *Journal of econometrics*, 208(2):487–506.
- Pieroni, L. and Salmasi, L. (2016). The effect of smoking habit changes on body weight: Evidence from the uk. *Economics & Human Biology*, 20:1–13.
- Rees, D. I. and Sabia, J. J. (2010). Body weight and smoking initiation: Evidence from add health. *Journal of Health Economics*, 29(5):774–777.
- Wehby, G. L. and Courtemanche, C. J. (2012). The heterogeneity of the cigarette price effect on body mass index. *Journal of health economics*, 31(5):719–729.

Figure 1: Trends in smoking and obesity (BRFSS 2002-2017)

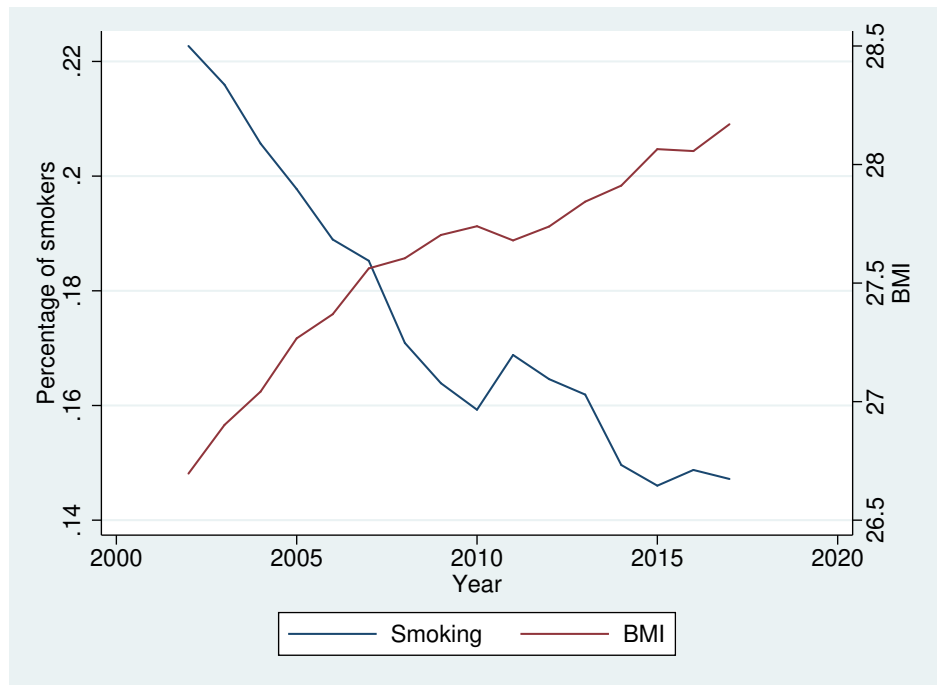


Table 1: Estimates from Previous Literature where BMI is Dependent Variable

| Study | Indep. Var. | Data | Method | Estimate | Interpretation |
|-------------------------------|------------------|---|----------|------------|---|
| Smoking Increases BMI | | | | | |
| Gruber and Frakes (2006) | Smoking Status | BRFSS 1984-2002 | 2SLS | 5.705 | Effect of smoking on BMI using cigarette taxes as an instrument |
| Gruber and Frakes (2006) | Cigarette Taxes | BRFSS 1984-2002 | OLS | -0.149 | Implied effect of a \$1 increase in cigarette taxes on BMI |
| Courtemanche (2009) | Cigarette Prices | BRFSS 1984-2005 | OLS | -0.234 | Effect of a \$1 increase in real price of cigarettes on BMI |
| Courtemanche (2009) | Cigarette Taxes | NLSY 1979 | OLS | -0.343 | Effect of a \$1 increase in real cigarette taxes on BMI |
| Wehby and Courtemanche (2012) | Cigarette Prices | BRFSS 1984-2005 | OLS | -0.28 | Effect of a \$1 increase in real price of cigarettes on mean BMI |
| Courtemanche et al. (2016) | Cigarette Prices | BRFSS 1990-2010 | OLS | -0.108 | Effect of a 1 s.d. increase in cigarette prices on BMI w/o including additional economic covariates |
| Smoking Decreases BMI | | | | | |
| Chou et al. (2004) | Cigarette Prices | BRFSS 1984-2002 | OLS | 0.486 | Effect of a \$1 increase in real price of cigarettes on BMI |
| Baum (2009) | Cigarette Taxes | NLSY 1979 | OLS | 0.328 | Effect of a \$1 increase in real cigarette taxes on BMI |
| Baum (2009) | Cigarette Prices | NLSY 1979 | OLS | 0.162 | Effect of a \$1 increase in real price of cigarettes on BMI |
| Fang et al. (2009) | Smoking Status | China Health and Nutrition Survey 2006 | 2SLS | -0.124 | Effect of a one-cigarette increase in smoking on BMI (instruments: community smoking rate, cigarette prices) |
| Liu et al. (2010) | Smoking Status | BRFSS 1998-2006 | 2SLS | -2.054 | Effect of workplace smoking bans on BMI |
| Courtemanche et al. (2016) | Cigarette Prices | BRFSS 1990-2010 | OLS | 0.036 | Effect of a 1 s.d. increase in cigarette prices on BMI after including additional economic covariates |
| Entezarkheir et al. (2016) | Cigarette Taxes | CCHS 2003, 2005 | OLS | 0.4 to 0.5 | Elasticity of obesity rate with respect to cigarette taxes |
| Pieroni and Salmasi (2016) | Smoking Status | BHPS 2004, 2006 | OLS | -0.59 | Effect of quitting smoking on BMI |
| Amialchuk et al. (2018) | Smoking Status | BHSIE 1996-2008 | 2SLS | -3.25 | Effect of quitting smoking on BMI for average Belarusian adult (instruments: cigarette prices, group-specific smoking rates) |
| Courtemanche et al. (2018) | Smoking Status | Lung Health Study | RCT/2SLS | -2.202 | Effect of smoking cessation treatment on BMI |

Notes: The implied -0.149 OLS reduced-form estimate of cigarette taxes on BMI from Gruber and Frakes (2006) is calculated by multiplying their 2SLS estimate by their first stage estimate of cigarette taxes on smoking status (-0.02611).

Table 2: BRFSS Summary Statistics, 2002-17

| | Non-Smoker | | Smoker | | Full Sample | |
|--------------------------|------------|--------|---------|--------|-------------|--------|
| | Mean | SD | Mean | SD | Mean | SD |
| BMI | 27.77 | 5.47 | 27.18 | 5.41 | 27.65 | 5.46 |
| Current Cigarette Smoker | - | - | - | - | 0.21 | 0.41 |
| Cigarette Tax (2017 \$) | 2.25 | 1.06 | 2.10 | 1.03 | 2.22 | 1.06 |
| Male | 0.43 | 0.49 | 0.46 | 0.50 | 0.43 | 0.50 |
| Black | 0.08 | 0.28 | 0.09 | 0.29 | 0.08 | 0.28 |
| White | 0.78 | 0.42 | 0.77 | 0.42 | 0.78 | 0.42 |
| Hispanic | 0.08 | 0.27 | 0.06 | 0.24 | 0.07 | 0.26 |
| Married | 0.63 | 0.48 | 0.43 | 0.49 | 0.59 | 0.49 |
| Divorced | 0.13 | 0.34 | 0.23 | 0.42 | 0.15 | 0.36 |
| Widowed | 0.03 | 0.17 | 0.05 | 0.21 | 0.03 | 0.18 |
| Grade 9-11 | 0.03 | 0.18 | 0.10 | 0.31 | 0.05 | 0.22 |
| High School | 0.24 | 0.43 | 0.38 | 0.49 | 0.27 | 0.44 |
| Technical Degree | 0.27 | 0.45 | 0.31 | 0.46 | 0.28 | 0.45 |
| College Graduate | 0.44 | 0.50 | 0.18 | 0.38 | 0.38 | 0.49 |
| Disposition Rate | 0.93 | 0.26 | 0.92 | 0.27 | 0.93 | 0.26 |
| Income (2017 \$) | 69,497 | 38,164 | 48,445 | 35,449 | 65,088 | 38,575 |
| State Unemployment Rate | 5.96 | 2.09 | 5.97 | 2.04 | 5.96 | 2.08 |
| Observations | 2,843,312 | | 753,331 | | 3,596,643 | |

Table 3: Variation in Nominal Cigarette Taxes (2002-2017)

| State | Tax in January 2002 (\$) | Tax in Dec 2017 (\$) | No. of increases |
|----------------------|--------------------------|----------------------|------------------|
| Alabama | 0.165 | 0.675 | 2 |
| Alaska | 1 | 2 | 3 |
| Arizona | 0.60 | 2 | 2 |
| Arkansas | 0.34 | 1.15 | 2 |
| California | 0.87 | 2.87 | 1 |
| Colorado | 0.20 | 0.84 | 1 |
| Connecticut | 0.50 | 4.35 | 8 |
| Delaware | 0.24 | 2.10 | 4 |
| District of Columbia | 0.65 | 2.50 | 2 |
| Florida | 0.339 | 1.339 | 1 |
| Georgia | 0.12 | 0.37 | 1 |
| Hawaii | 1 | 3.2 | 9 |
| Idaho | 0.28 | 0.57 | 1 |
| Illinois | 0.58 | 1.98 | 2 |
| Indiana | 0.155 | 0.995 | 2 |
| Iowa | 0.36 | 1.36 | 1 |
| Kansas | 0.24 | 1.29 | 3 |
| Kentucky | 0.03 | 0.60 | 3 |
| Louisiana | 0.24 | 1.08 | 3 |
| Maine | 1 | 2 | 1 |
| Maryland | 0.66 | 2 | 2 |
| Massachusetts | 0.76 | 3.51 | 3 |
| Michigan | 0.75 | 2 | 2 |
| Minnesota | 0.48 | 3.04 | 5 |
| Mississippi | 0.18 | 0.68 | 1 |
| Missouri | 0.17 | 0.17 | 0 |
| Montana | 0.18 | 1.70 | 2 |
| Nebraska | 0.34 | 0.64 | 1 |
| Nevada | 0.35 | 1.80 | 2 |
| New Hampshire | 0.52 | 1.78 | 5 |
| New Jersey | 0.80 | 2.70 | 5 |
| New Mexico | 0.21 | 1.66 | 2 |
| New York | 1.11 | 4.35 | 3 |
| North Carolina | 0.05 | 0.45 | 3 |
| North Dakota | 0.44 | 0.44 | 0 |
| Ohio | 0.24 | 1.60 | 3 |
| Oklahoma | 0.23 | 1.03 | 1 |
| Oregon | 0.68 | 1.32 | 4 |
| Pennsylvania | 0.31 | 2.60 | 4 |
| Rhode Island | 1 | 4.25 | 7 |
| South Carolina | 0.07 | 0.57 | 1 |
| South Dakota | 0.33 | 1.53 | 2 |
| Tennessee | 0.13 | 0.62 | 2 |
| Texas | 0.41 | 1.41 | 1 |
| Utah | 0.515 | 1.70 | 2 |
| Vermont | 0.44 | 3.08 | 7 |
| Virginia | 0.025 | 0.30 | 2 |
| Washington | 1.425 | 3.025 | 3 |
| West Virginia | 0.17 | 1.20 | 2 |
| Wisconsin | 0.77 | 2.52 | 2 |
| Wyoming | 0.12 | 0.60 | 1 |

Table 4: Effect of Smoking on BMI

| | (1) | (2) | (3) |
|--|----------------------------------|---------------------------------|---------------------------------|
| First Stage: Regress Smoking Status on Cigarette Taxes | | | |
| 2SLS | -0.005*** (0.002) [9.18] | -0.015*** (0.003) [20.77] | -0.006*** (0.001) [16.38] |
| 2-Step | -0.005*** (0.002) [128.16] | -0.018*** (0.000) [66.90] | -0.008** (0.000) [36.95] |
| NFI & CF | -0.006*** (0.002) [6.74] | -0.016*** (0.004) [20.48] | -0.006*** (0.002) [11.38] |
| Linear Reduced Form: Regress BMI on Cigarette Taxes | | | |
| Cigarette Taxes | -0.019 (0.050) | 0.012 (0.048) | -0.025 (0.042) |
| Second Stage: Regress BMI on Predicted Smoking Status | | | |
| OLS | -0.404*** (0.085) | -0.707*** (0.089) | -0.707*** (0.080) |
| 2SLS | 3.489 (8.260) | -0.792 (3.081) | 4.200 (6.302) |
| 2-Step | -0.678 (1.484) | -0.495 (0.937) | -1.059** (0.499) |
| NFI | -20.890 (16.168) | -4.526*** (1.248) | -6.210*** (0.631) |
| CF | -11.270 (8.615) | -5.482*** (1.510) | -7.366*** (0.760) |
| Observations | 3,596,643 | 3,596,643 | 3,596,643 |
| State FE | ✓ | | ✓ |
| Quarter-Year FE | ✓ | | ✓ |
| All covariates | | ✓ | ✓ |

Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. Note: the coefficient on the disposition variable in the first stage of the 2-step estimator is -0.021 (s.e. is 0.010). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effect of Smoking on BMI, Using Both Taxes and Disposition as Instruments

| VARIABLES | Linear Approach | | | Nonlinear Approaches | | |
|-----------------|-------------------------------|---------------------|----------------------|-------------------------------|----------------------|----------------------|
| | First Stage Smoking Status | Reduced Form BMI | 2SLS BMI | First Stage Smoking Status | NFI BMI | CF BMI |
| Cigarette Taxes | -0.006*** (0.001) | -0.026 (0.042) | | -0.006*** (0.002) | | |
| Disposition | -0.007** (0.003) | 0.389*** (0.024) | | -0.007** (0.003) | | |
| Smoker | | | -20.131** (9.643) | | -6.481*** (0.655) | -7.694*** (0.790) |
| Observations | 3,596,643 | 3,596,643 | 3,596,643 | 3,596,643 | 3,596,643 | 3,596,643 |

Standard errors are presented in parentheses and are robust to clustering at the state-level. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), and the control function approach (CF). All specifications include the full set of covariates and fixed effects. For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effect of Smoking on BMI by Age

| | OLS | 2SLS | 2-Step | NFI | CF |
|--------------------|---|-----------|-----------|-----------|-----------|
| Panel A: Age 18-40 | | | | | |
| | First Stage (DV = Smoking Status, Mean 0.224) | | | | |
| Cigarette Taxes | - | -0.006*** | -0.008*** | -0.006** | -0.006** |
| | - | (0.002) | (0.000) | (0.003) | (0.003) |
| | - | [8.62] | [75.27] | [5.17] | [5.17] |
| | Second Stage (DV = BMI, Mean 26.7) | | | | |
| Smoking Status | -0.195** | 4.039 | -4.217*** | -2.575*** | -3.276*** |
| | (0.082) | (9.465) | (0.542) | (0.539) | (0.676) |
| Observations | 1,218,758 | 1,218,758 | 1,218,758 | 1,218,758 | 1,218,758 |
| Panel B: Age 41-65 | | | | | |
| | First Stage (DV = Smoking Status, Mean 0.205) | | | | |
| Cigarette Taxes | - | -0.006*** | -0.011*** | -0.006*** | -0.006*** |
| | - | (0.001) | (0.000) | (0.001) | (0.001) |
| | - | [23.46] | [31.19] | [21.99] | [21.99] |
| | Second Stage (DV = BMI, Mean 28.1) | | | | |
| Smoking Status | -1.422*** | 4.112 | -0.784 | -6.058*** | -7.184*** |
| | (0.079) | (4.690) | (0.791) | (0.485) | (0.595) |
| Observations | 2,377,885 | 2,377,885 | 2,377,885 | 2,377,885 | 2,377,885 |

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * p<0.10, ** p<0.05, *** p<0.01.

Table 7: Effect of Smoking on BMI by Gender

| | OLS | 2SLS | 2-Step | NFI | CF |
|-----------------|---|-----------|-----------|-----------|-----------|
| Panel A: Male | | | | | |
| | First Stage (DV = Smoking Status, Mean 0.233) | | | | |
| Cigarette Taxes | - | -0.005*** | -0.006*** | -0.005*** | -0.005*** |
| | - | (0.002) | (0.000) | (0.002) | (0.002) |
| | - | [7.57] | [76.07] | [6.77] | [6.77] |
| | Second Stage (DV = BMI, Mean 27.8) | | | | |
| Smoking Status | -0.778*** | 3.871 | -6.893*** | -4.514*** | -4.844*** |
| | (0.091) | (5.632) | (0.499) | (0.532) | (0.579) |
| Observations | 1,564,310 | 1,564,310 | 1,564,310 | 1,564,310 | 1,564,310 |
| Panel B: Female | | | | | |
| | First Stage (DV = Smoking Status, Mean 0.194) | | | | |
| Cigarette Taxes | - | -0.007*** | -0.012*** | -0.007*** | -0.007*** |
| | - | (0.001) | (0.000) | (0.002) | (0.002) |
| | - | [24.56] | [25.25] | [14.58] | [14.58] |
| | Second Stage (DV = BMI, Mean 27.0) | | | | |
| Smoking Status | -0.521*** | 3.214 | 1.296 | -0.876** | -1.002* |
| | (0.078) | (6.941) | (0.976) | (0.392) | (0.509) |
| Observations | 2,032,333 | 2,032,333 | 2,032,333 | 2,032,333 | 2,032,333 |

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * p<0.10, ** p<0.05, *** p<0.01.

Table 8: Effect of Smoking on BMI: By Census Region

| | NE | MW | SO | WE |
|--|---------------------------------|---------------------------------|------------------------------|--------------------------------|
| First Stage: Regress Smoking Status on Cigarette Taxes | | | | |
| Mean of DV | 0.208 | 0.239 | 0.231 | 0.171 |
| 2SLS | -0.006*** (0.002) [13.31] | -0.009** (0.003) [7.85] | -0.009 (0.006) [2.77] | -0.000 (0.002) [0.00] |
| 2-Step | -0.007*** (0.000) [10.15] | -0.008*** (0.000) [24.43] | -0.013* (0.000) [4.99] | 0.002*** (0.000) [10.92] |
| NFI & CF | -0.007*** (0.002) [10.63] | -0.010*** (0.003) [9.17] | -0.011* (0.006) [3.50] | 0.001 (0.002) [0.39] |
| Second Stage: Regress BMI on Predicted Smoking Status | | | | |
| Mean of DV | 27.1 | 27.6 | 27.7 | 27.0 |
| OLS | -0.507*** (0.092) | -0.799*** (0.064) | -0.932*** (0.086) | -0.340* (0.160) |
| 2SLS | 15.345*** (1.914) | 1.233 (5.931) | 3.901 (4.456) | -3,394 (344,919) |
| 2-Step | -3.198*** (0.654) | -4.492*** (0.389) | 0.673 (0.577) | -0.900 (0.742) |
| NFI | -5.741*** (0.830) | -6.904*** (0.804) | -5.825*** (0.406) | -6.330*** (0.800) |
| CF | -6.521*** (0.969) | -7.680*** (0.926) | -7.111*** (0.544) | -7.180*** (0.892) |
| Observations | 696,936 | 879,807 | 1,122,084 | 897,816 |

All specifications include covariates, year-quarter fixed effects, and state fixed effects. Standard errors are presented in parentheses and are robust to clustering at the state-level. F / Chi-Square statistics are presented in brackets. Data are from the BRFSS 2002-2017. We weight observations by BRFSS-provided survey weights. Estimators include ordinary least squares (OLS), instrumental variables (2SLS), Newey correction to the forbidden regression (NFI), the control function approach (CF), and 2-step to correct for misreporting (2S). For comparison purposes, the first stage estimates for NFI and CF that we present are obtained by calculating the marginal effects. The first stage estimates for 2S that we present are obtained by regressing predicted true smoking status on cigarette taxes and controls via OLS. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.