The elasticity of corporate taxable income is a key policy parameter with limited evidence due to a lack of comprehensive data and viable variation. We provide crisp new evidence using the universe of C-corporations in the US and newly developed methods that exploit different types of variation. We find the elasticity of corporate taxable income is 0.64, substantially larger than previous estimates. In addition, we show that using newly developed methods are important in practice, because in their absence we would have underestimated the elasticity by 25%.

JEL:
Keywords:
Acknowledgements: We thank for helpful comments.
In fiscal year 2018, US corporate tax receipts fell by 31% to total $205 billion dollars, or 6% of total US tax receipts. At the same time, corporations faced the first reduction in the US corporate tax rate in decades. Although there was broad bipartisan support for a reduction in the US corporate tax schedule, which boasted the highest marginal tax rate in the OECD, there was widespread disagreement on what the optimal US corporate tax rate should be. On the one hand, the US competes for domestic investment on a global scale, which puts downward pressure on the optimal rate, and on the other hand the corporate tax base is an important receipts source for public expenditures, which pushes in the opposite direction.

In order to balance these and other concerns, we need to understand how corporations respond to the corporate tax rate. This behavior is summarized by the corporate elasticity of taxable income, which describes the sensitivity of firms to the corporate tax rate. To this end, neoclassical theory predicts that firms will decrease taxable income in response to higher tax rates, whether due to a policy change or across the marginal tax schedule. Saez (1999) underscores that theory predicts that firms will bunch at kink points in budget constraints and this bunching is evidence of sensitivity to tax rates.

More important than knowing that firms are sensitive to tax rates is knowing how sensitive they are. This question is an empirical question. Unfortunately, recent econometric work has shed light on a fatal flaw in most bunching estimation strategies. In particular, Blomquist and Newey (2017) and Bertanha, McCallum, and Seegert (2016) point out it is impossible to identify the magnitude of the ETI without additional assumptions about the underlying distribution of taxable income or additional variation.

Fortunately, the corporate setting is ideally suited to investigate the implications of the Blomquist and Newey (2017) and Bertanha et al. (2016) criticism for several reasons. First, as was pointed out by Patel, Seegert, and Smith (2016), the treatment of losses under the US corporate tax code exogenously shifts the marginal tax schedule in a firm-specific way. Because of this, Patel et al. (2016) show that firms bunch at points in the marginal tax schedule where the stock of loss carry-forwards is equal to kinks in the marginal tax
schedule. These bunching masses provide a unique setting with many kink points in which to test strengths and weaknesses of subsequently proposed methods. Second, Patel et al. (2016) provide an identification strategy for the bunching methodology that is based on a control group, overcoming the criticisms of Blomquist and Newey (2017). For this reason, the methodology of Patel et al. (2016) uniquely identifies the ETI and can therefore serve as a benchmark by which to compare other estimation strategies. Finally, following the framework of Bertanha et al. (2016), we evaluate the various parametric and semi-parametric distributional assumptions necessary for the implementation of bunching.

The range of available empirical estimates of the corporate elasticity of taxable income are sufficiently wide that they are uninformative. Previous attempts at estimating the corporate elasticity of taxable income have resulted in estimates ranging from 0.2 to 0.57 (Gruber and Rauh, 2007; Devereux, Liu, and Loretz, 2014; Patel et al., 2016). This range suggests that aggregate taxable income in FY2013 could have decreased by $2.3 billion to $10.4 billion in order to avoid a 1% increase in the marginal tax rate. Precision in the estimated magnitude of the CETI is important, and a transparent understanding of the benefits and costs of distributional assumptions is critical for effective public policy.

We undertake a careful evaluation of the estimated elasticity across thirty different firm specific kink points and seven different identification strategies. We find that there is support for a transparent assumption of underlying normality in the distribution of corporate firms, as is evidenced by the consistency we find in estimates across several methodologies. We conclude that the CETI is between 0.62 and 0.64 based on the best available estimation methods. This estimate is more than three times as large as the only prevailing estimate of the CETI based on US firms, and 12% larger than the most recent estimate based on UK firms.

We proceed as follows. In Section 1 we present a background on the elasticity of taxable income and bunching methodologies. In Section 2, we translate the framework of Bertanha et al. (2016) to the corporate setting. In Section 3.3 we undertake a series of Monte Carlo
simulations to better understand the expected behavior of the various bunching methodologies. In Section 4 we describe the corporate tax data, and in Section 5.2 we provide the best estimate of the CETI. Finally, in Section 6 we discuss the implications for our results.

1 Estimating Elasticities of Taxable Income

How individuals and firms respond to taxes occupies a central place in academic and policy analysis. There is a large labor literature focused on how labor supply changes in response to higher tax rates. This focus on labor is probably two-fold. First, labor supply is an important economic factor and taxes could create substantial deadweight loss if they are distorting labor supply. Fortunately, the near consensus is that taxes have little to no impact on the labor supply for prime-age males. Second, early data sources, such as the Panel Survey of Income Dynamics (PSID), collected data on labor supply.

The public finance literature has expanded the focus of distortions from taxes to capture all potential distortions. This, again, has probably due to two reasons. First, the efficiency costs of taxation when considering all potential distortions is likely to be larger than when focused solely on labor supply. Second, Feldstein (1999), among others, argued that the sum of distortions could be quantified by looking at how taxes distort taxable income. Since this point, the elasticity of taxable income (abbreviated ETI) has been the central parameter of interest in public finance.

Saez, Slemrod, and Giertz (2012) discusses the evolution of ETI methods estimates for the personal income tax—the bulk of work that has been done. It is telling that one of the main takeaways is, “researchers should be seeking better sources of identification,” (Saez et al., 2012). The main empirical hurdle to estimating ETI is a lack of adequate variation because i) tax rates rarely change, ii) when they do change it is likely due to economic conditions, iii) there rarely exists control groups that are unaffected by the tax change, and iv) governments are not likely to allow for field experiments where researchers are allowed to exogenously change tax rates for randomly selected individuals.
Due to these econometric hurdles, new methods exploiting different types of variation are quickly embraced. This is certainly true for the pioneering work by Saez (2010). In this paper, Saez demonstrates that tax payers bunch at kink points in the personal income tax schedule and that the amount of bunching is linked to the elasticity of taxable income. The importance of this paper cannot be overstated. Saez (2010) opened up the potential to estimate behavioral responses across many new dimensions because kink points in budget sets are ubiquitous. Following this paper, researchers sought out kinks and notches everywhere including real estate taxes (Kopczuk and Munroe, 2015), social insurance (Einav, Finkelstein, and Schrimpf, 2015), electricity prices (Ito, 2014), and labor regulations (Garicano, Lelarge, and Reenen, 2016).

Unfortunately, Blomquist and Newey (2017) and Bertanha et al. (2016) show that the classic bunching methods cannot identify the elasticity of taxable income without parametric assumptions. These papers prove that any positive elasticity is consistent with a given amount of bunching at a kink point. Said differently, the elasticities are completely driven by the parametric assumptions, which in classic methods are not transparent, cannot be tested, and as a result, often misspecified (Blomquist and Newey, 2017).

Fortunately, updated methods have been developed to capture the critical insight of Saez (2010) about bunching and use it to identify the elasticity of taxable income. Two early papers, Blomquist and Newey (2002) and Patel et al. (2016) develop methods to estimate elasticities using additional variation in budget constraints. Bertanha et al. (2016) provides a suite of empirical methods that include non-parametric bounds and semi-parametric and parametric estimates. These new methods are straightforward, transparent, easy to implement, and crucially identify the elasticity of taxable income.

We implement these new methods using a unique context to provide crisp new evidence on how firms respond to tax rates and to test in practice the advantages and disadvantages of these new methods. The prevailing elasticity of corporate taxable income estimates come from Gruber and Saez (2002) and Patel et al. (2016). The estimates differ substantially, and
as pointed out by Patel et al. (2016) these differences are probably due to different data sources and methods. Using the suite of new estimation techniques allows us to add precision to the elasticity estimates.

The following section develops a framework for estimating the corporate elasticity of taxable income and demonstrates how previous methods coincide and differ.

2 What Can We Learn From Bunching

We follow Bertanha et al. (2016) in framing bunching at kink points as a middle censored problem. This characterization formalizes the components necessary to identify the elasticity. The following section then demonstrates how different methods fulfill this set of components.

We build on Patel et al. (2016) in building a model of bunching for firms. The result of combining these insights is a model of bunching for firms built on firm optimization.

Firm $j$ begins period 1 with $X_j$ retained earnings and chooses how to change its capital stock, through investment and dividend policy. Firm $j$’s capital in period 2 is $K_j = X_j + I_j$, where investment, $I_j$, is the net of equity issuance and dividend payments. Firm $j$’s profits in period 2, net of costs such as depreciation, are a function of capital and its total factor productivity (TFP), $A_j > 0$,

$$
\Pi_j(K_j) = \left(\frac{1 + e}{e}\right)^{\frac{1}{1+e}} A_j^{\frac{1}{1+e}} K_j^{\frac{e}{1+e}}.
$$

(1)

Firms face a piece-wise linear budget set where we assume taxable income is profit, $Y_j(K_j) = \Pi(K_j)$.\footnote{See Patel et al. (2016) for a model where taxable income is profits net tax shields the firm uses. We abstract from those here for simplicity.} For simplicity, assume there is one kink point at $\eta$ in taxable income. Income below the kink point is taxed at $\tau_0 > 0$ and income above the kink point is taxed at $\tau_1 > \tau_0$.\footnote{We also assume $\tau_0 < 1$ and $\tau_1 < 1$.}

The firm maximizes value for its shareholders (there are no agency concerns in this model).
For expositional ease, we assume that the firm 2 liquidates after period 2. We define value as the value in period 2 net of period 1 capital. The firm discounts future payments at a rate \( r > 0 \), which is assumed to be the un-taxed rate of return of government bonds. Firm \( j \) chooses its period period 2 capital \( k_j \) to maximize shareholder value

\[
\max_{K_j} V_j = -rK_j + \tilde{Y}_j(K_j)
\]  

\[s.t. \quad \tilde{Y}_j = \mathbb{I}\{Y_j(K_j) < \eta\}[\eta - (1 - \tau_0)Y_j(K_j)]
+ \mathbb{I}\{Y_j(K_j) > \eta\}(1 - \tau_0)\eta + (1 - \tau_1)(Y_j(K_j) - \eta).
\]  

The optimal period 2 capital for firm \( j \) is found by solving the first-order condition and is given by

\[
K_j = \begin{cases} 
\frac{e}{1+e} \tilde{A}_j(1 - \tau_0)^{1+e} & , \text{if } 0 < \tilde{A}_j < \eta(1 - \tau_0)^{-e} \\
\frac{e}{1+e} \eta(1 - \tau_0) & , \text{if } \tilde{A}_j \in [\eta(1 - \tau_0)^{-e}, \eta(1 - \tau_1)^{-e}] \\
\frac{e}{1+e} \tilde{A}_j(1 - \tau_1)^{1+e} & , \text{if } 0 < \tilde{A}_j > \eta(1 - \tau_1)^{-e}.
\end{cases}
\]  

Most studies consider the elasticity of taxable income with respect to the net-of-tax rate, \((1 - \tau)\). Taxable income, substituting in the optimal capital in period 2 is,

\[
Y_j = \begin{cases} 
\tilde{A}_j(1 - \tau_0)^e & , \text{if } 0 < \tilde{A}_j < \eta(1 - \tau_0)^{-e} \\
\eta & , \text{if } \tilde{A}_j \in [\eta(1 - \tau_0)^{-e}, \eta(1 - \tau_1)^{-e}] \\
\tilde{A}_j(1 - \tau_1)^e & , \text{if } 0 < \tilde{A}_j > \eta(1 - \tau_1)^{-e}.
\end{cases}
\]  

The elasticity of taxable income with respect to the net-of-tax rate is \( \frac{\partial Y_j}{\partial (1 - \tau)} \frac{(1-\tau)}{Y_j} = e \). Taxable income is increasing in TFP, \( \tilde{A}_j \), until taxable income equals \( \eta \), where the tax rate increases
from $\tau_0$ to $\tau_1$. A set of firms with TFP in a certain range all report taxable income equal to $\eta$. This set of firms are often referred to as bunchers because the distribution of taxable income experiences a spike at the kink point. Then taxable income continues to increase with the TFP.

The seminal insight of Saez (2010) is that the amount of bunchers is increasing with the elasticity. Formally, the range of TFP where firms bunch is given by $\eta[(1 - \tau_1)^{-e} - (1 - \tau_0)^{-e}]$, which is increasing with respect to the elasticity $e$. Said differently, a bigger spike of bunching at a kink point is associated with a larger elasticity, *ceteris paribus*. As a first step, therefore, observing bunching is sufficient to sign the elasticity.

The second step is to quantify the magnitude of the elasticity given the amount of bunching at a kink point. Unfortunately, Bertanha, McCallum, and Seegert (2016) and Blomquist and Newey (2017) show that more structure is needed to identify the elasticity. Their insight is illuminated by formally stating the amount of bunching at a kink point,

$$B = \int_{\eta(1-\tau_0)^{-e}}^{\eta(1-\tau_1)^{-e}} f(\tilde{A})d\tilde{A},$$

where $f(\tilde{A})$ is the distribution of TFP. From this equation, it is clear that the amount of bunching at a kink point depends on the elasticity (in the bounds of the integral) and the distribution. Unfortunately, the distribution of TFP is unobserved, which means the amount of bunching at a kink point is not sufficient to identify an elasticity. Said differently, researchers must make additional assumptions or use additional variation to quantify the elasticity.

The following section discusses several methods for estimating the elasticity of taxable income.
3 How Can We Estimate Elasticities Using Bunching

3.1 Classic Methods

For comparison, we first estimate the elasticity of taxable income using two classic methods. First, we estimate the elasticity of taxable income using the method first outlined by Saez (2010). This method uses a trapezoid approximation to the unknown distribution. Bertanha, McCallum, and Seegert (2016) demonstrate this method may produce similar estimates—and estimates within their nonparametric bounds—when the trapezoid approximation is sufficiently close. Saez’s estimating equation is derived from equation (6) above,

\[
B = \int_{\eta(1-\tau_0) - \epsilon}^{\eta(1-\tau_1) - \epsilon} f(\tilde{A})d\tilde{A} \approx \Delta \tilde{A} \frac{f(\tilde{A}) + f(\tilde{A} + \Delta \tilde{A})}{2}
\]

\[
= \Delta \tilde{A} \frac{f(\tilde{A}) + f(\tilde{A} + \Delta \tilde{A})}{2} \]

\[
= \tilde{A} \left[ \left( \frac{1-\tau_0}{1-\tau_1} \right)^\epsilon - 1 \right] \frac{f(\tilde{A}) + f(\tilde{A} + \Delta \tilde{A})}{2} \] (7)

and is exactly equation (5) in Saez (2010).³

Second, we estimate the elasticity using the method described by Chetty, Friedman, Olsen, and Pistaferri (2011). This second method is the method most bunching papers follow (see e.g., Hansen, Miller, and Weber (2017) and Devereux et al. (2014)). This method uses a seventh-order polynomial to estimate the side limits of the distribution and a uniform distribution assumption. The mass of firms at the kink point when the distribution is assumed to be uniform can be written as,

\[
B = \int_{\eta(1-\tau_0) - \epsilon}^{\eta(1-\tau_1) - \epsilon} f(\tilde{A})d\tilde{A} = F(\eta(1-\tau_1) - \epsilon) - F(\eta(1-\tau_0) - \epsilon)
\]

\[
\approx f(\eta(1-\tau_0) - \epsilon) \left[ \eta(1-\tau_1)^{-\epsilon} - \eta(1-\tau_0)^{-\epsilon} \right].
\]

³Note that \( \frac{\Delta \tilde{A}}{\tilde{A}} = \left( \frac{1-\tau_0}{1-\tau_1} \right)^\epsilon - 1. \)
With this approximation, the change in taxable income can be written as,

$$\Delta Y = (1 - \tau_0)^e \left( \eta (1 - \tau_1)^{-e} \right) - F(\eta (1 - \tau_0)^{-e})$$

$$= \frac{(1 - \tau_0)^e}{f(\eta (1 - \tau_0)^{-e})} B.$$ 

Now, the elasticity can be written as a function of the mass of firms at the kink point, $B$, and observable factors,

$$e = \frac{\Delta Y}{\eta \ln \left( \frac{1 - \tau_0}{1 - \tau_1} \right)}$$

$$= \frac{b(\tau_1, \tau_2)}{\eta \ln \left( \frac{1 - \tau_0}{1 - \tau_1} \right)},$$

where, following Chetty et al. (2011), we define $b = Bg(\tilde{A})$. This estimation requires estimating $g(\tilde{A}) \equiv \frac{(1 - \tau_0)^e}{f(\eta (1 - \tau_0)^{-e})}$, which is the density of taxable income at the kink. In practice, this is done using a seventh order polynomial. This, however, is separate from the uniform distribution assumption, which allows the mass of bunching firms to be written as in equation (8).

Both of these methods are parametric (Blomquist and Newey, 2017). Blomquist and Newey (2017), unfortunately, demonstrate that there is no way of knowing or testing whether these parametric assumptions are consistent with the data.

### 3.2 New Methods

Given the impossibility of identifying the elasticity of taxable income from the amount of bunching at a kink point alone (Bertanha et al., 2016), several papers have proposed additional structure that does identify the elasticity. Blomquist and Newey (2002), the first to our knowledge, derives a nonparametric estimator using additional variation in the budget set. The clear advantage of this approach is that it does not require strict assumptions. The disadvantage, of course, is that a researcher must have the necessary additional variation.
The following subsections discuss several of these new approaches, which we use to estimate the corporate elasticity of taxable income. These methods include the control group method proposed by Patel et al. (2016) and the suite of options proposed by Bertanha, McCallum, and Seegert (2016). First, Bertanha, McCallum, and Seegert (2016) derive nonparametric bounds for $\eta$. This is a good first step that requires minimal assumptions. Second, they draw on the labor literature, which commonly estimates these models using maximum likelihood, to provide a semi-parametric MLE estimator. Third, they derive a flexible local estimator that is robust to other distortions in the distribution of taxable income. For example, the distribution of taxable income is often simultaneously subject to several piece-wise linear budget constraints thanks to the interaction of many tax and benefit schedules. Fourth, they derive a semi-parametric censored least absolute deviation estimator (CLAD).

### 3.2.1 Nonparametric Bounds (Bertanha, McCallum, and Seegert 2016)

The nonparametric bounds derived in Bertanha, McCallum, and Seegert (2016) provide a first step to estimating the elasticity of taxable income using bunching. In essence, the observed distribution of taxable income with a kink is distorted relative to the distribution in the absence of the kink. In particular, equation (5) demonstrates that the distribution of firms beyond the kink is pushed left; this causes some of the distribution to bunch at the kink. The key difficulty in estimating the elasticity comes from the fact that the researcher does not typically observe the distribution from where the bunchers came. The researcher is trying to push the bunched distribution back to the right, but with little guidance on how to do this properly.

The nonparametric bounds acknowledges that the distribution is unobserved but notes that the slope of the unobserved distribution is bounded above and below by some amount $M$. Given the bounds on the distribution, Bertanha, McCallum, and Seegert (2016) derive bounds on the elasticity. In essence, if the slope of the distribution is positive and steep then
the distribution of bunchers does not need to be pushed very far to the right. In this case, the furthest firm that moved to kink did not change their income by that much, and the elasticity will be small. In contrast, if the slope is negative and steep then the distribution will have to be pushed much further to the right; the furthest firm that moved to the kink changed their income by a lot, and the elasticity will be large.

We follow Bertanha, McCallum, and Seegert (2016) and calculate the bounds for a range of potential slopes $M$, according to the following equations,

$$
\xi_1 = 2 \left[ f(\eta^+)^2/2 + f(\eta^-)^2/2 + M B \right]^{1/2} - (f(\eta^+) + f(\eta^-)) \\
M (\ln(1 - \tau_0) - \ln(1 - \tau_1))
$$

$$
\xi_2 = -2 \left[ f(\eta^+)^2/2 + f(\eta^-)^2/2 - M B \right]^{1/2} + (f(\eta^+) + f(\eta^-)) \\
M (\ln(1 - \tau_0) - \ln(1 - \tau_1))
$$

3.2.2 Control Group Method (Patel, Seegert, and Smith 2014)

Patel et al. (2016) proposes a control group method that estimates the distribution in the absence of the kink using a set of firms that do not experience that kink. This method is straightforward, does not rely on distributional assumptions, and is applicable in many settings because piecewise linear budget sets often are shifted based on plausibly exogenous characteristics. In their setting, they use firms with slightly more net operating losses—which shifts the effective tax schedule for firms. The disadvantage is not all contexts will have a suitable control group.

The novelty of the control group method is how it estimates the change in taxable income. Remember, the elasticity is given by $e = \Delta Y / (\eta \ln(1 - \tau_0) / \ln(1 - \tau_1))$, where the only unobserved factor is $\Delta Y$. As we saw above, the classic methods estimate $\Delta Y$ by assuming either a quadratic approximation (Saez, 2010) or that the distribution is uniform (Chetty et al., 2011). Patel et al. (2016) estimate $\Delta Y$ using the following minimization problem,
\[ \min_{\Delta Y} \left( \sum_{\eta}^{\eta+1} h^e(y, n) - B(\tau_0, \tau_1) \right)^2 \]  

(12)

where \( h^e(y, \eta) \) is the number of firms in a bin around taxable income \( y \) with NOLs \( n \).

This approach, in essence, expands the distribution of taxable income right following the distribution of the control group, instead of assuming a slope as in the nonparametric bounds or a flat distribution as in the classic methods.

### 3.2.3 Middle Censoring Model (Burtless and Hausman 1978)

Taxable income given in equation (5) produces a distribution of taxable income that is characterized by a middle censored model. If we assume the distribution of TFP is normally distributed with unknown mean and variance, then we can estimate the elasticity using maximum likelihood and a generalization of a Tobit model that has been studied extensively (Burtless and Hausman, 1978). Bertanha, McCallum, and Seegert (2016) demonstrate how this model applies to the bunching case and derive the likelihood function of \( y_i \equiv \ln(Y_i) \)

\[
L(y_1, \ldots, y_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma} \phi \left( \frac{y_i - \text{eln}(1 - \tau_0) - \mu}{\sigma} \right)^{I\{y_i < \eta\}} \\
\times \left[ \Phi \left( \frac{\eta - \text{eln}(1 - \tau_1) - \mu}{\sigma} \right) - \Phi \left( \frac{\eta - \text{eln}(1 - \tau_0) - \mu}{\sigma} \right) \right]^{I\{y_i = \eta\}} \\
\times \frac{1}{\sigma} \phi \left( \frac{y_i - \text{eln}(1 - \tau_1) - \mu}{\sigma} \right)^{I\{\eta < y_i\}}. 
\]

(13)

This likelihood function can be extended to include covariates by assuming, \( \mu = x_i' \beta + \nu_i \). This loosens the assumption that TFP is normally distributed to the assumption that \( \nu_i \) conditional on \( x_i \) is normally distributed.

### 3.2.4 Flexibly Local Model (Berthana, McCallum, and Seegert 2016)

Berthana, McCallum, and Seegert (2016) demonstrate the normal distribution assumption in
the middle censoring model can be loosened by estimating a truncated model. The truncated model uses data only around the kink point and is more similar to the original bunching methods in spirit than the middle censoring model. The advantage of using data local to the kink point is that it abstracts from other distortions, such as kink points, that might exist due to other taxes and programs elsewhere in the distribution.

We follow Bertanha, McCallum, and Seegert (2016) and estimate the elasticity for a range of truncation from 10 percent to 100 percent. These estimates provide insights into other potential confounding factors that may be a threat to identification when assuming a globally normal distribution.

3.2.5 Other Methods

In addition, there are several papers that derive estimators for specific contexts. For example, Gelber, Jones, and Sacks (2017) focuses on earnings frictions that cause extensive decisions at retirement age and Einav, Finkelstein, and Schrimpf (2017) derive a structural model in the context of prescription drug insurance for the elderly in Medicare Part D. These are important advances and provide crisp new evidence in these contexts. We, unfortunately, cannot use these models in our setting.

Most of these methods focus on identifying the elasticity of taxable income in the absence of optimizing frictions (or measurement error). The identification of the elasticity in the presence of optimizing frictions, however, is an important topic and is the focus of Cattaneo, Jansson, Ma, and Slemrod (2018). In many contexts bunching is not isolated to a spike right at the kink point but is spread out as a hump. In these cases, Chetty et al. (2011) and Cattaneo et al. (2018) show that the elasticity estimate can be misidentified if optimizing frictions are not accounted for. The advantage of the method in Cattaneo et al. (2018) is that it can be applied to the data in a first step—to filter the data—and then one of the other methods can be used. In our setting, optimizing frictions is not our focus and the method by Cattaneo et al. (2018) is currently unavailable and, therefore, we either apply the
methods directly to the data or use the filtering process proposed by Bertanha, McCallum, and Seegert (2016).

### 3.3 Simulation Results

This section reports estimates using the methods described above on simulated data. The advantage of using simulated data is that we know the true elasticity. To test the sensitivity of different methods to their assumptions, we also vary the distributional assumptions. These results illuminate the advantages and disadvantages of different methods. The following section uses these methods with data on US corporations.

Table 1 reports our elasticity estimates in four different simulations. The first simulation assumes a normal distribution with a mean and standard deviation of 10.34 and 0.97, depicted in figure 1 Panel A. In this setting, all methods seem to perform well. Specifically, the control group method estimates an elasticity of 1.01, the MLE 1.02, the flexibly local 0.98–1.03, the classic 1.04, and the nonparametric bounds are (0.96, 1.27). The success of these methods suggests the added structure these methods impose were adequate and sufficiently close to the truth. For example, the MLE approach assumes correctly that the distribution is drawn from a normal distribution. Less obvious, however, is that the trapezoid assumption in the classic method would do well in this case.

The second simulation again assumes a normal distribution but shifts the distribution to have a mean and standard deviation of 4 and 0.97, depicted in figure 1 Panel B. Shifting the mean should have no affect on the control group method, the bounds, Tobit, or flexibly local estimates. The shift could, however, have an affect on the classic method if the kink point is now at a point such that a trapezoidal approximation does less well. This is exactly what we find. The estimates are all very similar to the estimates in the first row and to 1, the true elasticity, for all but the classic method which now has an estimate of 1.76. The large difference in the classic estimate suggests that it is sensitive to the parametric assumption it makes. Unfortunately, Blomquist and Newey (2017) demonstrate that there is no way of
knowing or testing whether this parametric assumption is reasonable or not.

The third simulation draws from an exponential distribution to test the sensitivity to different distributions, depicted in figure 2. The control group method and nonparametric bounds should not be sensitive to using a different distribution because they do not rely on distributional assumptions. The estimates using these methods are 1.01 and (0.85, 1.18), both close to the true elasticity of 1. The Tobit and flexibly local estimates assume the underlying distribution is normal and therefore could be sensitive to a different distribution. The estimates, however, are not very sensitive. The reason is the exponential distribution is in the linear exponential family and in the limit as the number of observations increase, the normal distribution assumption is reasonable.

The fourth simulation draws from a bimodal normal distribution that is not part of the linear exponential family, depicted in figure 3. This distribution is, however, normal locally. In this case, the Tobit performs poorly, with an estimate of 1.61, because the globally normal assumption is sufficiently poor. The flexibly local estimates, however, perform relatively well when less data is used. Panels A and B of figure 3 depict the estimated density using 75 and 50 percent of the data, respectively. From these graphs it is clear that the model fits the data better using 50 percent of the data. This can also be seen in the estimates, which are 1.48 and 0.98 when using 75 and 50 percent of the data, respectively.

Together these simulations suggest these methods may be complements rather than substitutes. Each of these methods provides advantages and disadvantages that are hard to test ex ante. The estimates from a set of these methods provides a well rounded understanding of the elasticity. In addition, researchers can remain agnostic about which set of estimates are most credible—leaving that to the reader to decide which assumptions are most likely to hold in the given context.

Different models may also be subtly estimating slightly different behavioral responses. At a larger level, most bunching estimators are thought to estimate local responses—e.g., excluding extensive margin responses. He, Peng, and Wang (2018) provides detailed analysis
comparing the classic bunching methods to differences-in-difference estimates. Using data from a Chinese tax change, He et al. (2018) finds a large difference in elasticity estimates between these methods.

4 Data on US Corporations

Our empirical analysis draws on administrative tax data that covers the universe of c-corporations from 2004 - 2015. During this time period, there were roughly 1.5 million c-corporations. Data are drawn from annual 1120 tax returns. We focus on measures related to tax liability, including net income (line 28), net operating loss deduction (line 29a), and taxable income (line 30). In addition, we gather measures of cash flow and size such as gross receipts (line 1c), total income (line 11), and total assets (schedule L line 15d). We focus on firms earning between -$100,000 and $100,000 in net income. This subset of the population data includes roughly 11 million firm-year observations.

Net income can be thought of as a measure of corporate profit in that it captures income less tax-deductible costs. For IRS purposes, “taxable income” is net income that is bounded below by zero. In addition, taxable income deviates from net income through the use of the net operating loss (NOL) deduction. The NOL deduction allows firms to utilize losses, or negative net income, earned in previous tax years to offset positive net income in the current tax year. During this period, firms earned an average of $4,380 in net income and held a stock of $6,871 in NOLs.

Prior to 2018, c-corporations faced a progressive marginal tax schedule with a top tax rate of 35% based on their taxable income. In any given year, the distribution of firms by taxable income has attributes that are consistent with an exponential or power law distribution. In particular, the large fraction of firms with negative net income are pushed into a mass at $0 in taxable income. In Figure 4, we see that as net income increases, the number of firms quickly decays symmetrically along a long tails.4

4A graph of taxable income would lump all firms with negative net income into the $0 bin, exacerbat-
We expect that firms should bunch at each of the statutory kink points in the marginal tax schedule. Indeed, we see evidence of this behavior even in the aggregate distributional graph shown in Figure 4. The size of the mass of firms bunched at these kinks is related to the magnitude of the elasticity of taxable income. Unfortunately, Blomquist and Selin (2010) and Bertanha et al. (2016) make clear that the magnitude of ETI cannot be identified without knowledge of the counter-factual distribution of firms that would be present in absence of the kink. In other words, while we are not surprised to see firms bunching at kinks in the marginal tax schedule, estimates of the magnitude of the ETI are intrinsically linked to distributional choices made by researchers.

In addition to firms amassed at kinks the marginal tax schedule, Patel et al. (2016) identify additional bunching masses in the distribution of c-corporations that are related to the stock of NOLs retained by firms. In particular, statutory kinks in the marginal tax schedule are shifted in a firm-specific way by the stock of NOLs carried by the firm. This feature of the corporate tax code provides a unique opportunity to study bunching across an infinite number of kink points. For example, in Figure 5, we see a spike in the distribution of firms with $10,200 in NOLs at $10,200 in taxable income. Generally, we expect that corporations bunch in a firm-specific way across all points where the corporate marginal tax schedule intersects with their stock of NOLs.

We bin our data based on net income and the stock of NOLs using $200 buckets. If we had only relied on kink points in the statutory marginal tax schedule, we would have expected two bunching points for firms with less than $100,000 in net income. The availability of firm-specific kinks expands the number of bunching points to 502 kink points, one at each point where net income is equal to the stock of NOLs. This dense experimental environment provides a unique opportunity to analyze the strengths and weaknesses of the various bunching methodologies.
5 Elasticity of Corporate Taxable Income Estimates

In this section we exploit unique features of the corporate setting and the methodological structure of Bertanha et al. (2016) to report a suite of estimates of the CETI. These estimates are based on transparent assumptions about the underlying distribution of firms and exploit different sources of variation. This suite of evidence allows the reader to choose the estimate that is most appropriate in light of the appropriateness of the assumptions necessary for identification. Stronger assumptions about the underlying distribution of firms will lead to more precise estimates but also risk sacrificing generalizability. In addition, differences in estimates across methods are informative on the appropriateness of distributional assumptions.

As with the unrestricted distribution of corporations, the distribution of firms conditional on their stock of NOLs does not appear to be drawn from a normal distribution. We have explored the importance of underlying normality for the success of each of the various bunching methodologies under more regulated Monte-Carlo simulations. We will use the control group methodology as a benchmark by which to compare the other estimation strategies. This will allow us to analyze the performance of the various parametric and semi-parametric distributional assumptions in the context of the shape and features of the corporate data.

Table 2 presents the estimated CETI according to seven different estimation methods. In Column (1), we implement the control-group methodology of Patel et al. (2016). In Column (2), we present the nonparametric theoretical bounds for the CETI. In Column (3), we estimate the CETI under the strongest parametric assumption of normality in the underlying distribution of firms. In Columns (4) - (6) we relax these assumptions, requiring normality only local to the kink, based on a middle censored fraction of the data. Finally, in Column (7) we implement the “Classic” method, based on Saez (2010).
5.1 CETI for Firms with $10k in NOLs

In Panel A of Table 2, we estimate the elasticity in the context of a singular firm-specific kink for firms with $10,000 in net income and $10,000 in NOLs. We start here in order to abstract away from any variation in estimates that might occur across kink points. In this context we first investigate the shape of the distribution in order to leverage insights from our Monte-Carlo simulations about how we expect the tested methods to perform relative to each other and the benchmark methodology.

To this end, Panel A of Figure 6 plots the full distribution of these firms and Panel B zooms in on the kink. Here we see that the overall distribution appears to be shaped according to power law dynamics on either side of $0, with a large mass of firms at $0 and a quick decay moving away from zero with long tails. Moreover, we can clearly see mass of firms bunching at $10k in net income, exactly where their taxable income starts to become positive. The Monte Carlo results of Section XX suggest that the Classic methods may perform particularly poorly if the trapezoid approximation is insufficient. On the other hand, if the distribution of corporate firms falls within the linear exponential family, we might expect both the Tobit MLE and the truncated normal assumption to perform well.

Panel A of Table 2 presents the range of CETI estimates for firms with $10k in NOLs. First, there is a consistency across estimates between the control-group benchmark, the Tobit MLE and the truncated normal. Based on the control-group methodology, we estimate the CETI to be 0.52, which is near the top of the theoretical bounds. This estimate is comparable to the parametric estimate of 0.58 based on a strong underlying assumption of normality. It is also comparable to the semi-parametric estimate of 0.53 based on a truncated normal distribution using 75% of the data. The tight range of these estimates lends support to the appropriateness of an underlying assumption of normality in the counter-factual distribution of firms absent the kink. This estimate range of (0.52, 0.58) lies near the upper bound of possible CETIs according to the BMS bounds and is substantially higher than the best previous estimate of the CETI for US firms, 0.2 (Gruber and Rauh, 2007). Moreover, given
support for the assumption of normality, we gain greater precision in the estimate of the CETI. Second, the estimate based on Classic methodologies is remarkably dissimilar from those based on normality, and it lies near the bottom of the theoretical bounds. If the Classic consistently appears as an outlier among the other estimates across all kink points, this would suggests that this assumption is inappropriate.

In Figure 7, we further investigate the appropriateness of the strong parametric assumption of underlying normality in distribution of firms. A visual inspection of the empirical and estimated density reveal significant overlap, particularly to the right of the kink. Moreover, Figure 8 relates the estimated elasticity to the semi-parametric truncated normal assumption. In this case, the estimated elasticity is fairly stable as long as a majority of data is used. Taken together, the consistency of the estimated CETI across several methodologies suggests that the true CETI for these firms is in the range of [0.52,0.58] for firms with $10k in NOLs.

5.2 CETI for Firms with $9k - $14k in NOLs

Next, we investigate the performance of the various estimation strategies across thirty different kink points between $9k and $14k. Figure 10 depicts the variation in the estimated ETI within this range. Here, we see that there is a negative relationship between the estimated elasticity and the firm-specific kink, and this trend is persistent across all estimation strategies. In addition we see that the relative effectiveness of each estimation method is largely consistent across all kink points. Said differently, the relationship between the various estimates seen at the $10,000 kink are roughly consistent across all 30 kink points.

Importantly, estimation via Classic methods consistently underestimates the CETI. This is consistent with findings in previous studies (Devereux et al., 2014). One potential reason for this is that the trapezoid approximation could be systematically biased. Unfortunately, as Blomquist and Newey (2017) note, there is no way of knowing or testing the appropriateness of the parametric assumptions in the classic method. On the other hand, the Tobit estimate, which assumes a normal distribution, follows the CETI estimated by the control-group method.
relatively well across all kink points. This gives support to hypothesis that the corporate
data may be distributed according to a distribution that is within the linear exponential family. In this case, the central limit theory supports the appropriateness of a normality assumption for a sufficiently large sample of data.

In order to summarize the elasticity estimate across these kink points, we report an average elasticity in Panel B of Table 2. To construct the average, each estimate is weighted by the the number of firms local to the kink point. The mean squared error summarizes the volatility of the distributional assumptions relative to the ETI identified by the control-group method. Consistent with what we saw in Figure 10, the average Tobit estimate is very close to the CETI identified by the control-group methodology, and the MSE of this methodology is small. On the other end of the spectrum, the Classic estimate is 40% smaller than the control-group method, and the MSE is large. Finally, we can see that the truncated normal assumption does worse as the data becomes more restricted, following a similar pattern as was seen at the $10k kink in Figure 8.

In practice, the wide variability, potential downward bias, and inability to test for the appropriateness of the underlying parametric assumptions supporting classic methods should serve as large deterrents. In the context of the corporate data, the evidence that we present here demonstrates a consistent similarity of estimates based on assumptions of normality and the benchmark estimate based on a control group methodology across thirty different kink points. This consistency suggests that the strong parametric assumption of normality underlying the Tobit method are supported by the data. By transparently moving to an estimation methodology that is supported by the underlying data, we are able add important precision to the estimate of the CETI. By aggregating estimates across 30 different kink points, we find that the plausible range of the CETI is (0.62, 0.64) based on the Tobit MLE and the control-group methodology.
6 Discussion

Understanding the behavioral response of corporations to changing marginal tax rates is of critical importance to public policy makers, especially in light of on-going efforts to overhaul the corporate tax code. Neoclassical theory predicts that firms will reduce taxable income to avoid higher tax rates, and empirical densities of corporations confirm that firms bunch near kink points in their marginal tax schedule. However, rigorous empirical estimates are required to identify the magnitude of this response. In order to undertake this empirical exercise, we employ 7 different estimation strategies across thirty kink points in the marginal tax schedule. By transparently investigating the relationship between the CETI and underlying distributional assumptions and by comparing these estimates against a well-identified benchmark, we are able to provide the best possible estimate of the sensitivity of firms to tax rates.

We find that the corporate data is well approximated by an underlying assumption of normality, and that the CETI lies in the range of (0.62, 0.64). This range of estimates is substantially higher than previous estimates of the CETI, which fall in the range of (0.2, 0.55) (Gruber and Rauh, 2007; Devereux et al., 2014; Patel et al., 2016). These estimates are based on observed bunching across thirty different kink points for c-corporations with a stock of NOLs between $9k and $14k.

To quantify these results, we consider the implications of the impact on taxable income of a 1% decrease in the marginal tax rate between 2004 and 2013. The sheer width of plausible estimates, between $38 and $71 billion fewer dollars in taxable income, underscores the importance of precision in the estimated CETI. Our results suggest that firms would have reduced taxable income by an amount near the top of this range: between $67 and $69 billion in lost taxable income in order to avoid a small 1% increase in the marginal tax rate. This translates to roughly $14 billion fewer dollars in corporate tax receipts, enough money to fund agencies like the Food and Drug Administration, which had a discretionary budget of
just $4 billion, for several years.

As important as understanding the magnitude of the CETI is understanding the mechanisms that allow for a $67-$69 billion decrease in taxable income. Firms might intertemporally reclassify income or engage in other reporting mechanisms, or they may shift or reduce planning real activity such as hiring or investment. These mechanisms are associated with a societal dead-weight loss, and precise estimates of the magnitude of the CETI are the first step towards quantifying the dead-weight loss.

7 Conclusion

Corporate tax policy remains in the political spot light in the US because of its implications for investment, innovation, employment, and wage growth. Despite the importance of corporate tax policy relatively little is known about how sensitive firms are to tax rates. Two factors contribute to this lack of empirical evidence. One, comprehensive data on corporate taxable income has been unavailable and two, there have been relatively few corporate tax rate changes in the US. We provide new crisp evidence using IRS administrative data and newly developed methods that exploit new and different types of variation.

We find the elasticity of corporate taxable income is 0.52, which suggests firms are more sensitive to tax rates than previous estimates suggest. This estimate is fairly consistent using a wide variety of methods, each relying on different identification assumptions. This lends confidence to the estimate.

We also find that new methods that exploit bunching at kink points perform substantially better than previous methods. We show that in simulations classic methods can be substantially off, providing elasticity estimates up to 76% different from the true elasticity estimate. In the corporate data we find that classic methods seem to be systematically downward biased, a finding similar to other papers (Devereux et al., 2014). In practice, the classic method produced estimates that were 25% smaller than all other methods. This demonstrates that using these new methods is critical to identifying the true elasticity.
Our analysis also suggests future research should present a suite of estimates. Our simulation results demonstrate that different methods will perform better in different scenarios. Specifically, methods that make strong distributional assumptions, such as global normality, can add precision if this assumption is true but can perform poorly if it is not. Reporting a suite of estimates ranging from strong to weak allows the reader to decide the most appropriate assumption for the given context. In addition, differences in estimates provides insights into the appropriateness of different assumptions.
References


Saez, E. (1999). Do taxpayers bunch at kink points?


This table reports the estimates of the elasticity of taxable income using four different simulated data sets described in the right panel. The true elasticity is 1 in all data sets. The first two simulations use a normal distribution. The third simulation uses an exponential distribution and the fourth uses a bimodal normal distribution. The columns report the estimates using the control group method of Patel et al. (2016), the nonparametric bounds from Bertanha et al. (2016), a Tobit estimated with maximum likelihood, a locally flexible truncated distribution using 75, 50, and 25 percent of the data, and a classic method.

<table>
<thead>
<tr>
<th></th>
<th>PSS Control</th>
<th>BMS Bounds</th>
<th>Tobit MLE</th>
<th>Trunc 75</th>
<th>Trunc 50</th>
<th>Trunc 25</th>
<th>Classic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\varepsilon = 1)$</td>
<td>1.01 (0.96, 1.27)</td>
<td>1.02 1.03 0.98 1.02 1.01 1.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\varepsilon = 1)$</td>
<td>1.02 (0.88, 1.17)</td>
<td>1.03 1.01 1.01 1.01 1.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exponential</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\varepsilon = 1)$</td>
<td>1.01 (0.85, 1.18)</td>
<td>1.01 0.99 1.00 1.00 0.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bimodal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\varepsilon = 1)$</td>
<td>1.01 (0.88, 1.21)</td>
<td>1.61 1.48 0.98 0.94 0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Obs.</th>
<th>Mean</th>
<th>Std.</th>
<th>Kink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal $(\varepsilon = 1)$</td>
<td>8 m</td>
<td>10.34</td>
<td>0.97</td>
<td>7</td>
</tr>
<tr>
<td>Normal $(\varepsilon = 1)$</td>
<td>8 m</td>
<td>4</td>
<td>0.97</td>
<td>7</td>
</tr>
<tr>
<td>Exponential $(\varepsilon = 1)$</td>
<td>8 m</td>
<td>9.9</td>
<td>9.9</td>
<td>7</td>
</tr>
<tr>
<td>Bimodal $(\varepsilon = 1)$</td>
<td>8 m</td>
<td>16.5</td>
<td>15.5</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 2: Corporate Elasticity of Taxable Income

This table reports the estimates of the elasticity of taxable income based on administrative tax data from 2004 - 2014 for c-corporations with net income between -$100,000 and $100,000. Firms are grouped into $200 bins based on net income and the stock of NOLs. The columns report the estimates using the control group method of Patel et al. (2016), the nonparametric bounds from Bertanha et al. (2016), a Tobit estimated with maximum likelihood, a locally flexible truncated distribution using 75, 50, and 25 percent of the data, and a classic method.

<table>
<thead>
<tr>
<th></th>
<th>(1) PSS BMS</th>
<th>(2) Tobit</th>
<th>(3) BMS Trunc</th>
<th>(4) BMS Trunc</th>
<th>(5) BMS Trunc</th>
<th>(6) BMS Trunc</th>
<th>(7) Classic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Firms with NOLs = $10k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CETI (ε)</td>
<td>0.52 (0.33,0.62)</td>
<td>0.58</td>
<td>0.53</td>
<td>0.46</td>
<td>0.39</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.08</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Panel B: Firms with NOLs = $9k - $14k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CETI (ε)</td>
<td>0.64 (0.35,0.66)</td>
<td>0.62</td>
<td>0.57</td>
<td>0.50</td>
<td>0.44</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Panel C: Implications of 1% ↓ in τ 2004 - 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income Δ</td>
<td>69 (38,71)</td>
<td>67</td>
<td>61</td>
<td>54</td>
<td>47</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>
Panel A: Simulation Normal Distribution I

Panel B: Simulation Normal Distribution II

Figure 1: Simulation Normal Distributions
Panel A: Exponential Distribution

Panel B: Zoomed In

Figure 2: Simulation Exponential Distribution
Figure 3: Simulation Bimodal Distribution Locally Flexible Estimates
Figure 4: Distribution of Firms with $0 in NOL stock
Figure 5: Firm Specific Kink Points
Panel A: Full Distribution

Panel B: Zoom On Kink=$10k

Figure 6: Distribution Of Corporations: NOL=$10k
Figure 7: Kink = $10k : Normal Assumption

Figure 8: Range of ETI for Kink = $10k: Truncated Normal Assumption
Figure 9: Kink = $10k : 50% Trimmed

Figure 10: Corporate Elasticity of Taxable Income: Firm-Specific Kinks ($9k - $14k)
Figure 11: Kink = $10K : BMS Bounds