Can Fiscal Rules Constrain the Size of Government? An Analysis of the "Crown Jewel" of Tax and Expenditure Limitations

Paul Eliason
Department of Economics
Duke University

Byron Lutz
Research Division
Federal Reserve Board of Governors

December 2017

We are grateful to Carol Hedges, Director of the Colorado Fiscal Institute, for a helpful discussion concerning TABOR in general and the treatment of the TABOR surpluses in the Census of Governments data in particular. Natalie Mullis of the Colorado Legislative Council provided both valuable insight into TABOR and numerous useful documents and data. Phyllis Resnick provided valuable conversation on TABOR and local governments. We thank Chris Hansen and Moshe Buckinsky for very helpful conversations concerning the use of the synthetic control method to simultaneously examine two outcomes. We thank numerous conference and seminar participants for helpful comments; in particular, we thank Patrick Button, Therese McGuire, and Shanna Rose for high quality feedback as conference discussants. Lena Yemelyanov provided outstanding assistance with the computational challenges of the county-based synthetic cohort estimates and Katherine Richard and Kelsey Moran provided excellent research assistance. Paul Eliason thanks various funding sources within Duke University for support.

Can Fiscal Rules Constrain the Size of Government? An Analysis of the "Crown Jewel" of Tax and Expenditure Limitations

Fiscal rules attempt to alter budget outcomes by constraining policy makers. They have been one of the primary responses to the recent spat of fiscal policy failures around the globe—e.g. Greece, Puerto Rico and Detroit. It is unclear, though, whether these rules cause a change in budget outcomes, are evaded by policy makers, or merely ratify the existing preferences of a jurisdiction's voters and officials. We ask if fiscal rules are capable of altering budget outcomes by examining what is arguably the most stringent set of fiscal rules in the U.S.—Colorado's Taxpayer Bill of Rights (TABOR). TABOR applies to all sub-national levels of government in Colorado, sets tight caps on essentially all forms of government revenue and in theory has almost no escape clauses which would allow officials to violate the caps. Previous examinations of TABOR have universally come to the conclusion that it significantly reduced both taxation and spending – i.e. that it caused a reduction in the size of government. To evaluate TABOR, we explore several ways in which the synthetic control methodology of Abadie et al. (2010) can accommodate multiple outcome variables (taxes and expenditures). We settle upon a novel approach of estimating treatment effects for multiple outcomes simultaneously. Although there will always be a degree of uncertainty over external validity when a policy is enacted in only a single state, our results provide no evidence that TABOR affected the level of taxes or spending in Colorado and are precise enough to rule out large negative effects. Numerous robustness checks buttress this conclusion; in particular robustness to alternative estimation strategies is demonstrated. In sum, no support is found for the contention that fiscal rules alter budget outcomes. Instead, TABOR appears to have been partly evaded by policy makers and voters despite its stringency and partly nothing more than a ratification of the state's preference over the size of its public sector.

Paul Eliason
Department of Economics
Duke University
paul.j.eliason@gmail.com

Byron Lutz
Research Division
Federal Reserve Board of Governors
Stop # 83
20th and C Streets, NW
Washington, DC 20551-0001
byron.f.lutz@frb.gov

Representative democracies often produce poor fiscal outcomes such as large and persistent deficits. Such outcomes may reflect structural deficiencies in fiscal institutions. For instance, deficits may arise as a result of the common-pool problem in which the costs of deficits are widely dispersed, but the benefits of deficit-financed spending are highly concentrated. Another example involves asymmetric information between voters and officials. It can be costly for voters to monitor complex budget processes and this may allow officials to spend in excess of voters' preferences.

The chief response to problems of this type has been the introduction of fiscal rules which aim to alter budget outcomes by constraining policy makers. Examples include the budget frameworks adopted by the U.S. Congress (e.g. Gramm-Rudman-Hollings), numerical budget targets and non-partisan budget agencies in the European Union, balanced-budget rules and super-majority requirements in U.S. states and tax and expenditure limitations for both state and local governments in the U.S. In the wake of the recent string of fiscal crises around the globe – the debt crisis in southern Europe, the downgrading of U.S. government debt over deficit concerns, Puerto Rico's inability to pay its debt, and municipal bankruptcies in the U.S. – fiscal rules are likely to take on ever greater importance. Indeed, the EU significantly tightened it budget rules under the Fiscal Compact of 2012.

There are two broad schools of thought concerning fiscal rules (Poterba, 1997). The "public choice" view holds that budget rules are important constraints on political actors and causally alter budget outcomes. In contrast, the "institutional irrelevance view" holds that political actors systematically evade the intent of the rules while adhering to their letter. The rules are therefore seen as nothing more than a "veil" which can be easily pierced by political actors. Finally, there is a third possibility, closely related to the institutional irrelevance view: Budget rules may simply fail to bind. For instance, tax and expenditure limitations may express preference for small government. If elected officials make tax and spending decisions in line with these preferences regardless of whether or not a fiscal rule is in place, then a limit will not *cause* a change in the size of government, but merely ratify an existing preference over the size of government.

In this paper we ask if fiscal rules are capable of altering budget outcomes. We focus our attention on tax and expenditure limitations (TELs) – fiscal rules widely applied to both state and local governments in the U.S. These rules attempt to address the principal-agent problem between voters and elected officials over the proper size of government.[1]

Poterba and von Hagen (1999) note that empirical investigations of fiscal rules typically suffer from an important methodological tension. On one hand, econometric based studies can offer sound statistical properties, but rarely account for the institutional richness of fiscal rules. On the other hand, case studies allow for considerable nuance but "defy statistical analysis." In this paper we bridge these two approaches using the synthetic control method of Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015). The method allows us to hone in on the most prominent TEL in the U.S. – Colorado's Taxpayer Bill of Rights (TABOR) – in detail, while simultaneously providing precise quantitative inference with which to assess the statistical robustness of our conclusions.

TABOR is fertile ground for investigating the efficacy of fiscal rules because it is widely considered the most stringent TEL in the U.S. Put more colorfully, TABOR is "the crown jewel of the tax limitation movement" (Poulson, 2005b) and places Colorado on "the nation's strictest fiscal diet" (Bridges, 2004). Intense debate surrounds TABOR. Some contend that it appropriately restrains the size of Colorado government by resolving the principal-agent problem: "TABOR replaces ambiguous fiscal contracts between citizens and politicians with an explicit contract" (Poulson, 2005a). Furthermore, it has been argued that TABOR boosts economic growth.[2] Others believe it reduces the quality of public services in the state and unnecessarily constrains policy makers (e.g. Hedges 2003 and Lav and Williams 2010). For instance, many contend that it has reduced funding to discretionary portions of the state budget (e.g. higher education and public health) while having little effect on areas whose costs

[1]Of the set of budget rules in use in the U.S., TELs are the most directly aimed at restraining growth in the size of governments. Most of the other fiscal rules in use in the U.S., such as balanced budget requirements and debt limitations, primarily aim to achieve budgetary balance.

[2]E.g. New (2010a). For a dissenting view see McGuire and Rueben (2006).

are driven by factors outside the budget process (e.g. Medicaid and corrections). Moreover, TABOR may reduce the state's ability to respond to shifting economic conditions (e.g. James and Wallis 2004; Frates 2005).

Despite the acrimonious debate, there is *universal* agreement among all observers that TABOR reduced the size of government in Colorado. Appendix Table A1 contains a literature review of publications concerning TABOR and a selective review of the numerous policy pieces on the limit. Every item in the Appendix either provides evidence that TABOR reduces the size of government, cites other sources in support of this claim, or simply asserts the claim. The press has presented a similar view: One of the country's leading pundits has repeatedly extolled TABOR for constraining the size of government (Will, 2005, 2011), as has the editorial board of one of the nation's most prominent newspapers (Wall Street Journal, 2002, 2004). On the other side of the debate, the editorial board of another leading national newspaper has criticized TABOR as "anti-tax zealotry" which has prevented Colorado from adequately funding its public services (New York Times, 2015).

An evaluation of TABOR involves a number of challenging methodological concerns surrounding the fact that only a single state has ever enacted the policy. We attempt to surmount these difficulties by using the aforementioned synthetic control methodology to construct a synthetic Colorado from a weighted combination of states other than Colorado. The weights are chosen so that both taxes and spending in the synthetic Colorado mimic the behavior of these outcomes in the actual Colorado in the period before TABOR was enacted. The path of taxes and expenditures in the synthetic Colorado after TABOR's enactment then provides a counterfactual for what would have occurred in Colorado in the absence of TABOR.

The synthetic control methodology has a number of attractive features and is "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens, 2017). In relation to this study, the method has three primary advantages. First, although selecting an appropriate control group is of great importance in any policy evaluation, this process is often ad hoc and arbitrary. The synthetic control method provides a formal, data-driven method for choosing the control group and evaluating its appropriateness. Second, large sample inference methods are typically inappropriate when a control group is comprised of only a few units. The synthetic control method overcomes this difficulty by executing placebo tests on all states other than Colorado and assessing the prevalence of false positives. Third, the methodology is quite general in how it controls for unobservable factors that influence the common time trend of the treatment and control groups. Notably, it is more general than a fixed-effect estimator because it allows the influence of fixed, unobservable characteristics to have a *time-varying* influence on the outcome. This becomes a distinct advantage if, as suggested by Waisanen (2010), the unobserved preference for small government manifests itself differently over the course of the business cycle.

Unlike past uses of the synthetic control method, our analysis requires the examination of two outcome variables – taxes and expenditures. We therefore explore ways in which the synthetic cohort methodology can accommodate multiple outcomes of interest. Our preferred, and novel, approach is to simultaneously estimate treatment effects for taxes and expenditures with a single synthetic Colorado.

Despite the advantages of the synthetic control approach, there are inherent limitations to studying a policy which has only been enacted in a single state. In particular, we acknowledge that both estimating effects over a long period of time and establishing external validity is challenging; a level of uncertainty will remain, particularly with regards to external validity. On the other hand, TABOR is uniquely stringent in design and therefore provides a unique opportunity for evaluating the efficacy of stringent fiscal rules. Moreover, we employ a number of robustness checks in an attempt to mitigate the limitations of having only a single treated state. In particular, we perform a much more disaggregated, county-based analysis in the spirit of Kline and Moretti's (2014) evaluation of the Tennessee Valley Authority (TVA). We also conduct a regression-based dynamic difference-in-difference evaluation to demonstrate robustness to an alternative empirical methodology and to address a principal shortcoming of the synthetic control approach – its inability to control for time-varying factors in the post-TABOR period. Notably, we document the robustness of our conclusions to controlling in an extremely flexible manner for non-TABOR fiscal rules.

Our results provide no evidence that TABOR influenced any budget outcome in Colorado. In particular, the path of taxes in post-TABOR Colorado is nearly identical to that of taxes in the synthetically constructed Colorado suggesting TABOR did not suppress tax collections. Our numerous robustness checks buttress this conclusion. Moreover, our results are precise enough to rule out large negative effects.

Our reading of the qualitative evidence suggests that three factors produced this outcome. First, TABOR appears to have been partly nothing more than a sign, or ratification, of the state's preferences over the size of government. TABOR thus appears to be a solution to a perceived principal-agent problem which in large part does not exist. Second, it reflects limited evasion of the policy's intent by policy makers. Third, it reflects buyers remorse, as voters relaxed, but did not repeal, the policy on several occasions.

The conclusion that TABOR did not effect budget outcomes is important for at least three reasons. First, there have been attempts to enact TABOR-style TELs in at least 20 states since 2004 (Lav, 2009). Given the strong claims and counter-claims surrounding the limit and the possibility that multiple states could adopt similar arrangements, our findings have significant policy relevance. They suggest that neither the hopes of the proponents of TABOR-like policies, nor the fears of their opponents, are likely to be realized. Second, they have broader implications for TELs in general. Given that TABOR is the most stringent TEL in the nation and was drafted explicitly to address the perceived inadequacies of previous TELs, it seems unlikely that less stringent TELs are meeting their objective of constraining government. Finally, and most significantly, our results have implications with regards to fiscal rules in general. Our results fail to lend support to the public choice view that such institutions can improve or alter budgetary outcomes or restrain the size of government. Instead they are most consistent with the institutional irrelevance view of fiscal rules.

The remainder of the paper is organized as follows. Section 1 provides background on tax and expenditure limits in general and TABOR specifically. Section 2 discusses the details of our extension of the synthetic control method. Section 3 discusses the data. Section 4 presents our empirical findings and Section 5 concludes.

1 Tax and Expenditure Limitations and Colorado's TABOR

1.1 Tax and Expenditure Limitations

State-level tax and expenditure limitations seeking to restrain broad categories of spending and taxation are a relatively modern phenomenon dating from the late 1970s. New Jersey passed the first such limit in 1976. By 1982 17 states had enacted limits. Another wave of TEL passage occurred in the early 1990s and currently 30 states utilize this type of fiscal rule (Waisanen, 2010). At the local level, 46 states have a TEL which applies to at least one type of government such as cities, counties, school districts, etc. (ACIR, 1995). Many of these restrict property taxation (Anderson, 2006).

The existing literature on the effect of state-level TELs on budget outcomes is mixed and inconclusive–see Kousser et al. (2008a) for a review. The inconclusiveness may partly reflect failure to address the endogeneity of TEL passage.[3] Several papers have attempted to address this endogeneity, however, and the results remain mixed. For instance, Rueben (1997) uses the availability of direct legislation and recall procedures to instrument for the limits and finds that TELs reduce growth in state government expenditures. Kousser et al. (2008a,b) use a variety of techniques to address the endogeneity of TELS, including an event-study research design, and conclude that Colorado's TABOR is the only TEL to constrain total state and local government spending. The evidence on local government TELs is more conclusive and suggest that these limits have been successful at restraining growth in government (e.g. Poterba and Rueben 1995; Cutler et al. 1999; Dye et al. 2005; Brooks et al. 2012).

[3]As noted by Rueben (1997), states with voters desiring small government may be more likely to enact a TEL. Alternatively, states with unusually brisk growth in government may be more likely to pass a TEL. Either scenario would produce a spurious correlation between TELs and budget outcomes.

1.2 TABOR

Colorado has a long history with tax and expenditure limits.[4] The state's first tax limit, which applied to property taxes, was passed by the legislature in 1913. The first expenditure limitation, which applied to state General Fund expenditures, was enacted by the legislature in 1977. In 1991, the year prior to TABOR's passage, the legislature enacted a new, more restrictive General Fund limitation.

Voters in Colorado can directly amend the state constitution through an initiative procedure. The first attempt to enact a TEL through this process occurred in 1966. The measure failed, as did attempts in 1972, 1976 and 1978. Pre-cursors to TABOR—comprehensive limits intended to significantly restrict the scope of government in the state—were presented to voters in 1986, 1988 and 1990, but failed to pass. Modifications were made to these proposals and in 1992 TABOR became part of the Colorado Constitution after being approved by nearly 54 percent of voters.

TABOR is extremely stringent. It applies to *all* sub-national governments in Colorado including the state government, municipalities and school districts. It has three primary provisions: a requirement of voter approval for all tax increases, a revenue growth limit and a prohibition on revoking existing spending limits without voter approval.

The first provision, voter approval for tax increases, is broad and applies to tax rate increases, property tax millage increases, increases in property tax assessment ratios, extension of expiring taxes, the introduction of new taxes and any tax policy change which results in a net increase in revenue.[5] Tax decreases, on the other hand, do not require approval.

The second provision, the revenue cap, limits the annual increase in revenue. The provision is also quite broad, as it applies to almost all taxes and fees leveled in the state. At the state government level, the cap applies not only to the General Fund, but also to most cash

[4]This subsection draws heavily from Baron et al. (2003). It also relies on James and Wallis (2004), Colorado Fiscal Policy Institute (2004) and Martell and Teske (2007).

[5]For example, in 2012 Colorado became the first state to legalize the personal consumption of marijuana. Because of TABOR, the new marijuana industry could not be taxed without voter approval. Voters approved such taxation in 2013.

funds held by the state including the Unemployment Insurance Fund, the Transportation Fund and the Higher Education Fund. Thus, for example, an increase in tuition at public colleges counts toward the state's overall permissible revenue increase. Several categories of revenue are exempted, though, including grants from the federal government, enterprise revenue and pension fund revenue. At the state-level, revenue growth is constrained to equal population growth plus inflation (defined as the Denver-Boulder-Greeley Consumer Price Index). For school districts, the limit is equal to inflation plus growth in student enrollment. For local governments other than school districts, the limit is equal to inflation plus the net growth in taxable property due to construction.

The third provision, voter approval to revoke spending limits, prevents statutory expenditure limits from being removed by elected officials. In 1991, the state legislature capped the annual increase in General Fund appropriations to 6 percent. Prior to TABOR, this restriction could have been adjusted or eliminated by the legislature. Under TABOR, though, it can only be changed with voter approval.

TABOR is widely considered the most stringent TEL in the U.S.[6] In the words of a proponent, TABOR is "America's best and most effective revenue limit"(New and Slivinski, 2005). Several factors account for this perceived stringency. First, because it is a constitutional amendment it can only be amended or repealed by citizen vote. Furthermore, it contains almost no escape clauses. In general, only a natural disaster permits the legislature to deviate from TABOR even temporarily. Second, TABOR applies to all levels of government. A natural mechanism for evading a TEL is to shift tax and spending responsibilities to the non-limited level of government (e.g. Skidmore, 1999; Baron et al., 2003; Blom-Hansen et al., 2014; Zhang, 2017). TABOR prevents such actions. Third, the definition of revenue is unusually broad and includes not only taxes but almost all revenue raised by state and local governments. E.g., the revenue limit cannot be subverted by replacing tax revenue with user fees. Fourth, the revenue growth rate limit at the state-level – population growth plus

[6] See for example McGuire and Rueben (2006); Waisanen (2010); Resnick (2002).

inflation – is more restrictive than the more commonly used limit of personal income growth. If income growth outstrips inflation, revenue as a share of income will decline. The revenue limits at the local level are similarly extremely restrictive. Fifth, until 2006 the revenue limit was based on the previous year's actual revenue, not the previous year's limit. Thus, a decrease in revenue caused by an economic downturn will cause a permanent ratcheting down of revenue in all future years.

The state government did not breach the revenue cap until the 1997 fiscal year (which began on July 1, 1996), the fourth year that TABOR was in effect. It then exceeded the revenue limit for five consecutive years until the 2001 recession pushed revenues below the cap in the 2002 fiscal year. Almost \$3.25 billion in surplus revenue was collected over this period, the bulk of which was returned to tax payers through tax credits introduced into the state income tax system. See appendix section 6.1 for additional details on the TABOR rebates.

Rebates have also been issued by local governments (Hedges, 2003). The amount of these varies as a function of the underlying growth in local tax bases (primarily the property tax but also the sales tax) as well as the growth factor specified in the limit – enrollment growth for school districts and the value of new construction for municipalities, counties and special districts.[7] Overrides of the TABOR revenue cap have been common at the local level: Through 2011, 523 overrides have appeared on local ballots and 87% of these have passed. In addition, there has been 669 votes to raise local tax rates, with 55% passing (Colorado Municipal League, 2012).

TABOR has been modified by voters twice since its initial passage. Amendment 23, enacted in 2000, mandates that state expenditures for K-12 education must increase annually by 1 percent above inflation. In addition, a portion of income tax collections were diverted into an education fund exempt from the TABOR revenue limit: Taxpayers in essence chose

[7]Discussions with multiple Colorado officials suggest that no systematic data is available on these rebates. In most cases, the limits were adhered to not by issuing rebates, but by lowering the millage rate of the property tax in order to avoid violating the revenue caps (Greenwood and Brown, 2000).

to forgo a portion of their TABOR refund in order to increase education spending. The amendment also contained the implication that a higher share of the state budget would go to K-12 education over time.[8]

The interaction of TABOR and the 2001 recession would have produced a *permanent* reduction in the size of Colorado government had voters not intervened. Government revenue fell sharply following the 2001 recession. As the subsequent economic recovery got underway, government revenue began to rebound. However, because the TABOR revenue limit is based on the previous year's revenue, a permanent ratcheting down of revenue was slated to occur. In the absence of TABOR, this revenue growth would have been available to restore funding to programs cut during the fiscal crisis. Under TABOR these funds had to be rebated to taxpayers.

In November of 2005 voters again amended TABOR by passing Referendum C (Watkins, 2003). The referendum had broad support, including many business leaders, the Republican governor and the leadership of the Democratic legislature. Referendum C revokes the TABOR revenue limit for 5 years and sets a new limit after this "timeout period" has passed. The new limit is based on the previous year's revenue cap instead of the previous year's actual revenue (and continues to adjust for inflation and population growth). Thus, the referendum removes the ratchet effect. In the first year after the timeout (fiscal year 2011), the limit is set to the highest annual revenue over the 5-year period (plus inflation and population growth). Revenue in excess of the TABOR limit during the 5-year period is earmarked to health care, education, pension plans and transportation.

Significantly, Referendum C did not revoke the requirement of voter approval for tax increases nor the prohibition on revoking spending limits without voter approval. Furthermore, policy actors were likely to be influenced by the knowledge that TABOR would again bind revenues in the future. Thus, although TABOR was less stringent during the timeout period, it remained in effect and may have influenced budget decisions in the state.

[8] The discussion of Amendment 23 is a simplification which captures the key elements. See Legislative Council Staff (2001) for additional information.

Finally, it is important to note that TABOR is stringent with regard to constraints on policymakers—i.e. elected officials. It is does not seek to bind the actions or preferences of future voters; quite the opposite, TABOR explicitly includes provisions to allow future voters to avoid fiscal outcomes they view as undesirable (e.g. the availability of citizen tax override votes). More fundamentally, there is no mechanism available by which TABOR could bind future voters at the state-level — future voters will always have the ability to adjust a fiscal rule such as TABOR (Kousser et al., 2008a).[9] Thus, we view the fiscal effects of Amendment 23 and Referendum C as part of the overall fiscal effect of TABOR.[10]

2 Empirical Approach

The synthetic control method of Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015) brings increased rigor to comparative case studies. Such studies compare the evolution of an outcome in a single "treated" unit to the evolution of the outcome in a "control group" of unaffected units. Prominent examples in economics include comparing outcomes in Miami at the time of the Mariel Boat Lift to outcomes in other southern U.S. cities (Card, 1990) and comparing employment in New Jersey at the time of an increase in the minimum wage to employment in Pennsylvania (Card and Krueger, 1994). [11] More recent examples of case studies include Kline and Moretti's (2014) evaluation of the TVA, Bohn et al.'s 2014 study of the effect of a policy in Arizona intended to reduce the employment of undocumented immigrants (Bohn et al., 2014), and Dustmann et al.'s 2017 examination of an immigration shock in Germany (the last two studies both employ the synthetic control method.)

Case studies typically suffer from two drawbacks – ambiguity over choice of control group and lack of an appropriate inference method. The synthetic control method provides a

[9]Analogously, future legislatures will always have the ability to adjust a fiscal rule put into place by current legislature.

[10]Stated differently, the efficacy of a fiscal rule in causally altering budget outcomes is partly a function of the future behavior of the set of agents with the ability to change the rule. Viewed in this light, Amendment 23 and Referendum C are endogenous responses to TABOR whose effect is properly incorporated into the overall TABOR treatment effect.

[11]The effect of the Mariel Boat Lift was recently re-examined by Peri and Yasenov (2015) using the synthetic control method.

solution to both problems: It uses a formal, data-driven procedure to select a control group and uses a series of falsification checks to provide inference.[12] Despite these advantages of the methodology, our case study—like those case studies cited immediately above—remain subject to certain limitations including an unavoidable degree of uncertainty over the external validity.

We extend the synthetic control approach to a setting where there are two outcomes of interest—taxes and expenditures—both of which may be affected by the treatment. If there is a desire to compare estimated treatment effects across outcomes, or draw conclusions which rely jointly on both estimates, then estimating the procedure separately for each outcome may be problematic because it will involve drawing two distinct control groups – one for each outcome. Differences in estimates based on differing control groups will conflate differences in the behavioral response to the policy (e.g. the policy induced policy makers to alter expenditures but not taxes) and differences in the composition of the control group. We address this issue by adapting the synthetic control approach to draw a single control group appropriate for two outcomes simultaneously. In the following section, we lay out the standard synthetic control approach. In section 2.2 we extend this standard approach to the two-outcome case.

2.1 Treatment Effect

Assume initially that there is a single fiscal outcome under study: tax collections. We define the TABOR treatment effect for state i in year t as

$$\alpha_{i,t} = Y^1_{i,t} - Y^0_{i,t}, \tag{1}$$

[12]The method was developed to study the influence of terrorism on the economy of Spain's Basque region (Abadie and Gardeazabal, 2003) and was subsequently used to assess the effect of a tobacco control program in California (Abadie et al., 2010). Other applications include, but are not limited to, estimating the value of political connections (Acemoglu et al., 2016), the effects of California's affirmative action ban on college racial integration (Hinrichs, 2012), the effect of natural disasters on economic growth Cavallo et al. (2013), and the effects on local economic activity of nuclear power plants in Japan (Ando, 2015).

where $Y^1_{i,t}$ are total taxes per capita collected by state and local governments under TABOR and $Y^0_{i,t}$ are collections in the absence of TABOR.

The classic obstacle to attaining $\alpha_{i,t}$ is the inherent unobservability of $Y^0_{i,t}$: We cannot observe what tax collections would have been in Colorado had TABOR not been enacted. The task at hand is to produce a credible estimate of $Y^0_{i,t}$. We do so by forming a weighted average of tax collections from a group of potential control states $i = 1, 2...., J$, called the donor-pool, which are unaffected by TABOR: $\sum_{i=1}^{J} w_i^* Y_{i,t}$ where w_i^* are the weights on the states in the donor pool. We refer to this estimate of $Y^0_{i,t}$ as the tax collections in a "synthetic Colorado."

The treatment effect estimate is simply the difference in collections between Colorado and the synthetic Colorado

$$\hat{\alpha}_{J+1,t} = Y_{J+1,t} - \sum_{i=1}^{J} w_i^* Y_{i,t} \tag{2}$$

where Colorado is indexed as state $J + 1$. The estimator rests on the assumption that, in the absence of TABOR, tax collections in Colorado would have equaled collections in the synthetic Colorado.

Calculating $\hat{\alpha}_{J+1,t}$ requires selecting the donor-pool weights w_i^*. Formally, consider a set of state weights $W = (w_1, w_2, ..., w_J)$ where $w_i \geq 0$ for $i = 1, 2, ..., J$ and $\sum_{i=1}^{J} w_i = 1$. The optimal W satisfies

$$W^* = \underset{W}{\operatorname{argmin}} (X_1 - X_0 W)' V (X_1 - X_0 W) \tag{3}$$

where X_1 is a (k x 1) vector of pre-TABOR variables from Colorado which have predictive power for tax collections and X_0 is a (k x J) matrix of the same pre-TABOR predictors from the donor pool states.[13] V is a diagonal matrix whose values are weights designating the relative importance of each predictor for the outcome variable (i.e. tax collections). The weights obtained from solving equation (3), $W^*(V^*) = (w_1^*, w_2^*, ..., w_J^*)$, determine the composition of the synthetic Colorado.

[13]For example, we include the state unemployment rate in our predictor vector because it is likely predictive of demand for certain public services such as Medicaid.

Since W^* is a function of V, the choice of V is crucial. V is chosen to satisfy

$$V^* = \underset{V}{\operatorname{argmin}}(Z_1 - Z_0 W^*(V))'(Z_1 - Z_0 W^*(V)) \tag{4}$$

where Z_1 is a (T_o x 1) vector containing total taxes per capita in Colorado in each of the T_o years of the pre-TABOR period and Z_0 is a (T_o x J) matrix of vectors containing the same pre-TABOR variable for the J states in the donor-pool.

Intuitively, equation (3) ensures that the synthetic Colorado is as similar as possible to the actual Colorado in terms of the observable fiscal predictors X. Equation (4) weights the importance of these fiscal predictors such that tax collections in the pre-TABOR synthetic Colorado mimic collections in pre-TABOR Colorado as closely as possible.

The procedure is more nuanced than simply generating a synthetic Colorado which provides the best unconditional match for Colorado in the pre-treatment period. Instead, it produces as close a match as permitted through the intermediation of the fiscal predictors. Stated differently, the method aims to produce a pre-TABOR match which reflects fundamental drivers of fiscal outcomes.[14] Relative to the simplistic "best-fit" procedure, the use of the fiscal predictors increases the credibility of the assumption that, in the absence of TABOR, the synthetic and actual Colorados would have had the same tax collections.

Finally, a significant advantage of the synthetic control method is its ability to account for unobservable factors that influence the common time trend of the treatment and control groups. To see this, first define the outcome under study for any state i as: $Y_{i,t} = Y^0_{i,t} + \alpha_{i,t} D_{i,t}$, where $D_{i,t}$ is an indicator for treatment. Abadie et al. (2010) demonstrates that $Y^0_{i,t}$ can be characterized as a factor model: $Y^0_{i,t} = \delta_t + \Theta_t Z_i + \lambda_t \mu_i + \varepsilon_{i,t}$. Notice that the generality of this form allows the unobservable confounders, μ_i, to have a time-varying influence (i.e. λ is allowed to vary with t). Alternative methods for estimating equation (1), such as difference-

[14]For example, the business cycle differential affects different sectors of the economy. A state's tax base, and hence a state's tax collections, will therefore fluctuate over time as a function of the state's sectoral mix. In order for the tax collections in the synthetic Colorado to provide a plausible counterfactual for the tax collections in the real Colorado, it is desirable for the synthetic Colorado's sectoral mix to be similar to Colorado's sectoral mix.

in-difference or fixed-effect methods, restrict the effect of the unobservable confounders to be time-invariant.

The generality of the synthetic control method is a substantial virtue in the context of TABOR. TELs have a strong tendency to be enacted during economic downturns (Waisanen, 2010). This can be seen in Figure A-1. The initial burst of TEL passage occurred during the economic turbulence of the late 1970s and early 1980s and there was a second bout of activity around the time of the early 1990s recession. In contrast, there was much less activity during the middle and later portions of both the 1980s and 1990s. It therefore seems likely that the unobserved preference for small government manifests itself differently over the course of the business cycle. The synthetic control method is capable of accounting for this likelihood, whereas less-general methods are not.

2.2 Multiple Treatment Effects

As TABOR targeted both taxes and expenditures it is natural to estimate a treatment effect for both outcomes. The examination of two outcomes involves an important methodological choice between three possible estimation strategies. The first strategy is to re-estimate the procedure for each outcome. This approach is straightforward and produces a superior pre-treatment fit between the actual and synthetic states for each outcome. The disadvantage is that the control group, or synthetic state, changes for each outcome. We will refer to this method as the "floating weight" method.

The second strategy is to find an optimal set of state weights for one outcome (e.g. taxes) and impose the same state weights on the secondary outcome (e.g. expenditures). This eliminates the undesirable changing of the control group across outcomes. However, this approach has at least two drawbacks. First, the secondary outcome in the synthetic state may not closely match the secondary outcome in the actual state in pre-treatment period. If so, the synthetic control group will fail to provide a compelling counterfactual for the secondary outcome. Second, the procedure requires prioritizing one outcome over the other. We will refer to this method as the "priority weight" method as it requires prioritizing one outcome over the other.

In order to overcome the drawbacks inherent in the floating and priority weight approaches, we develop a third option in which the synthetic control procedure is allowed to optimize over both outcomes simultaneously, as intermediated by the fiscal predictors, to create a single set of donor-pool weights which represent a plausible counterfactual for both outcomes. This strategy utilizes a single synthetic control group, does not require prioritizing one outcome over the other, and potentially yields a close match between the actual and synthetic states in the pre-treatment period for both outcomes. We refer to this method as the "simultaneous weight" method.

Suppose we are interested in $\alpha_{i,t}$, the effect of TABOR on tax collections, and $\gamma_{i,t}$, the effect of TABOR on expenditures:

$$\begin{bmatrix} \alpha_{i,t} = Y^1_{i,t} - Y^0_{i,t} \\ \gamma_{i,t} = E^1_{i,t} - E^0_{i,t} \end{bmatrix} \tag{5}$$

where $E^1_{i,t}$ represents state and local expenditures under TABOR and $E^0_{i,t}$ represents expenditures in the absence of TABOR.[15] We seek a single set of weights that can be used to estimate both $\alpha_{i,t}$ and $\gamma_{i,t}$ by taking the difference between Colorado and the synthetic Colorado

$$\begin{bmatrix} \hat{\alpha}_{i,t} = Y_{J+1,t} - \sum_{i=1}^{J} w_i^* Y_{i,t} \\ \hat{\gamma}_{i,t} = E_{J+1,t} - \sum_{i=1}^{J} w_i^* E_{i,t} \end{bmatrix} \tag{6}$$

In order to select this set of donor-pool weights, we modify equation (4) slightly to include both outcomes:

$$V^* = \underset{V}{\operatorname{argmin}} (\tilde{Z}_1 - \tilde{Z}_0 \tilde{W}^*(V))'(\tilde{Z}_1 - \tilde{Z}_0 \tilde{W}^*(V)) \tag{7}$$

where $\tilde{Z}_1$ is a ($2T_o$ x 1) vector containing a (T_o x 1) vector of total taxes per capita in

[15] Although our presentation here uses 2 outcomes, the approach generalizes to n outcomes.

Colorado in each of the T_o pre-TABOR years stacked on top of a corresponding (T_o x 1) vector of expenditures per capita in Colorado. $\tilde{Z}_0$ is a corresponding matrix of vectors containing the same pre-TABOR variables for the J states in the donor-pool. The tax and expenditure variables are standardized by transforming them into z-scores such that they have a mean equal to zero and a standard deviation equal to one. Absent this standardization, the procedure would potentially place more implicit weight on one outcome than the other.[16]

The intuition behind the estimator is little changed relative to the single outcome case. Equation (3) continues to require the synthetic Colorado to be as similar as possible to the actual Colorado in terms of the observables X. Equation (7) weights the importance of the predictors X such that *both* taxes and expenditures in the pre-TABOR synthetic Colorado mimic their counterparts in the actual pre-TABOR Colorado as closely as possible.

2.3 Inference

The synthetic control estimate rests on the assumption that, in the absence of TABOR, tax collections in Colorado would have evolved as they did in the synthetic TABOR. The inference procedure developed by Abadie et al. (2010) assess the credibility of this assumption. A TABOR treatment effect is generated for each donor-pool state using the synthetic control method. These estimates are referred to as placebo treatment effects. If the synthetic control method produces accurate counterfactual tax collections and expenditures, the placebo treatment effects should be small, relative to Colorado, given that these state were not treated by TABOR. As a result, the Colorado treatment effects are credible only if they are sufficiently large relative to the distribution of the placebo treatment effects (i.e. taxes or expenditures).

We can calculate a significance level (p-value) for the one-sided test of the hypothesis that TABOR did not reduce tax collections using the method suggested by Cavallo et al. (2013) where $p-value_t = \frac{\sum_{i=1}^{J} \mathbb{1}(\hat{\alpha}_{i,t}^{pl} < \hat{\alpha}_{J+1,t})}{\#\ of\ Placebo\ States} = \frac{\sum_{i=1}^{J} \mathbb{1}(\hat{\alpha}_{i,t}^{pl} < \hat{\alpha}_{J+1,t})}{J}$, where $\hat{\alpha}_{j,t}^{pl}$ is the estimated treatment

[16] However, if one outcome is judged more important than the other outcome, the procedure could be weighted so as to prioritize the more important outcome.

effect on tax collections of placebo state j in year t. The p-value ranks the Colorado treatment effect within the distribution of the placebo treatment effects. Using the same logic, we can construct a lower confidence limit to establish the magnitude of tax reduction we have the statistical power to detect. If the estimated effect of TABOR falls below this limit, we can reject the null hypothesis. Failure to do so does not allow us to rule out negative TABOR effects of all possible magnitudes, but only magnitudes beneath the lower confidence limit.

2.4 The Donor Pool

The states in the donor pool include both states with and without state-level TELs. Our estimates therefore capture the effect of TABOR against a counterfactual of not having a TABOR-like policy. By design, the counterfactual explicitly includes the possibility of having an ordinary state-level TEL. This is appropriate and quite natural for two reasons. First, *Colorado had an ordinary TEL in place at the time of TABOR's enactment* and it is therefore extremely natural to include states with these ordinary TELs in the control group. Specifically, Colorado's ordinary TEL was enacted in 1977 and then strengthened in 1991. Colorado is fairly typical in this regard as 30 states have a standard TEL in place and these were all enacted in the late-1970s or later. Second, TABOR is widely considered a "super-TEL" that is substantially more effective than other TELs; indeed TABOR was explicitly designed to remedy the perceived ineffectiveness of standard TELs. Thus, it is appropriate to test the efficacy of this "super-TEL" relative to ordinary TELs. Overall, we believe that the question of "what are the effects of TABOR relative to all non-TABOR states of the world" is both a well posed question and a policy relevant one.

That said, to properly interpret our results it is quite important to assess the extent to which they reflect binding TELS—and stringent fiscal rules more broadly—in the control states. The synthetic control research design, through its implicit use of a double difference, nets out the effect of ordinary TELs in the control group as long as they are in place for the duration of the sample period.[17] As a result, only TELs which are enacted within sample

[17]The double difference is [post-Tabor Colorado taxes - post-Tabor synthetic Colorado taxes] - [pre-Tabor

will influence our results. This is a relevant issue as our synthetic Colorado contains several states which enacted fiscal rules within the sample period. We therefore marshal evidence from three approaches to shed light on the extent to which our TABOR treatment effect reflects fiscal rules in the control states: our standard synthetic control approach, the county-based synthetic control approach, and the regression-based dynamic difference-in-differences approach. Viewed jointly, these pieces of evidence suggest that non-TABOR fiscal rules in the control states do not explain our TABOR treatment effect estimates.

Finally, we omit Alaska and Wyoming from the donor pool because they display extreme tax revenue volatility owing to the taxation of petroleum extraction. (However, including these states in the donor pool has essentially no effect on the results.) Thus, there are 47 states in the donor pool.

3 Data

Most of the data we use come from the US Census Bureau. Budget outcome variables are from the State and Local Government Finances Database. All budget outcomes are used on a real inflation adjusted per-capita basis and capture the activity of both state and local governments. TABOR applies to all levels of (sub-national) government and can only be fully evaluated with data on both state and local government outcomes. Our sample of budget outcomes runs from 1977 to 2012. We extend our sample to 2012 as some readers may be interested in the long-run effects of the policy. However, this is an unusually long window over which to evaluate a policy and the assumption that the synthetic Colorado provides a valid counterfactual for the real Colorado becomes gradually less plausible with time from TABOR's implementation. The estimates in the out years, particularly those from the Great Recession and afterwards, should be interpreted with a degree of caution.

There are two important aspects of the Government Finances Database which affect our analysis. First, local government finance statistics for 2001 and 2003 are unavailable. While

Colorado taxes - pre-Tabor synthetic Colorado taxes]). Any effect of the ordinary TEL on tax collections in the synthetic Colorado is therefore differenced out.

this does not influence the composition of the synthetic Colorado (which is based solely on data from the pre-TABOR period), it does mean we cannot include these years in our analysis. In the graphs in Section 4, the 2001 and 2003 outcomes are linear interpolations provided solely to ease visual inspection of the graphs. Second, the Colorado tax collections reported to the Census Bureau fail to account for the TABOR rebates and therefore *significantly overstate* actual collections.[18] We net out the TABOR rebates from the Census tax collections data using information contained in the Annual Reports of the State of Colorado Department of Revenue and administrative data provided to us by the Colorado Legislative Council (which is part of the Colorado state government). Failure to do so would strongly bias us against concluding that TABOR reduced tax collections.[19] [20]

Age and education data are from the Decennial Census (1970, 1980, 1990 and 2000). Ideally we would have these data at a yearly frequency. However, because the synthetic control method uses the mean of the predictors over the pre-treatment period, the fact that we only have these at a 10-year frequency does not pose a significant problem. State population estimates are from the Decennial Census and the Intercensal Population Estimates. Data on the sectoral composition of each state's economy are from the Standard Industrial Classification, published by the BEA. State unemployment rates are from the Local Area Unemployment Statistics published by the BLS. Data describing party political control of the legislature and governorship are from the Partisan Division of American State Governments database,

[18] Although the majority of TABOR rebates were returned to taxpayers as a "sales tax" rebate, the rebates were administered through the state income tax system and were based on a taxpayer's federal tax form adjusted gross income (AGI). Labeling the rebates as "sales tax" increased the amount of state income tax that could be deducted from federal tax liability. Had the rebates been an income tax rebate, it would have lowered an individual's state income tax liability and would therefore have lowered the federal deduction. Apparently as a result, the income tax data reported to the Census Bureau did not net out the TABOR rebates even though the rebates significantly lowered actual income tax collections. Nor were they netted out of sales tax collections.

[19] The failure of the Census data to account for the rebates is mentioned in Hedges (2003). We thank Carol Hedges of the Colorado Fiscal Policy Institute for assistance in understanding the issue. We also thank Natalie Mullis of the Colorado Legislative Council for assistance with this issue and for providing the administrative data with which we adjust the Census data.

[20] The ability to address issues of this type is a significant advantage of the comparative case study method relative to studies which use policy changes in many states (or other units). The researcher would rarely become aware of critical issues of this nature while conducting such a study.

published by the Inter-University Consortium for Political and Social Research (ICPSR). We rely on the "Fiscal Survey of States" (fall edition, fiscal years 1996 through 2001) produced by the National Association of State Budget Officers (NASBO) for data on changes in state tax policy.

4 Results

4.1 Simultaneous Weight Results

We first present our preferred results produced using the the simultaneous weight approach. The first column of Panel A, Table 1 displays the mean value of the fiscal predictors X for Colorado during the pre-TABOR period of 1977 to 1991. (All major categories of variables included in the X vector are displayed, although not all specific variables are displayed. For instance, we include ten sectoral mix variables, but only display three.) The next two columns display the mean values for the synthetic Colorado and the pool of 47 donor states, respectively. The synthetic Colorado is a substantially better match for Colorado than is the donor pool. In particular, the synthetic Colorado is a good match in terms of census division and political variables. Notably, it matches Colorado's preference over this period for Republican control of both houses of the legislature with a Democratic governor. It is also an overall better match in terms of the sectoral mix of the state economy, the unemployment rate, the percent of population which is elderly and educational attainment. The synthetic Colorado is a somewhat worse match than the donor pool along a couple of dimensions, though, such as youth population share. It can be inferred that the synthetic matching procedure places less importance on these state attributes.

As displayed in Table 2, which contains the W weights, the synthetic Colorado is constructed from six donor pool states. Four of these states, Arizona, Kansas, Nevada and Utah are quite similar to Colorado in terms of region and the political composition of their state governments. These states comprise 86% of the synthetic Colorado. New Hampshire and New York make up the remainder.

Panel A of Figure 1 plots the per-capita tax collections of both Colorado and the synthetic

Colorado.[21] The synthetic collections are simply the weighted average of collections from the six states forming the synthetic Colorado. These synthetic collections are nearly identical to the Colorado collections in the pre-TABOR period. The tightness of this match strongly suggests, but does not prove, that the synthetic Colorado is an appropriate counterfactual for Colorado. The fit in the post-TABOR period is nearly as tight (other than in a few years in the mid-2000s), suggesting that TABOR had absolutely no effect on tax collections. The close correspondence between Colorado and its synthetic counterpart in the post-treatment period is remarkable given that this period is "out of sample" as the synthetic Colorado is constructed without any information from this period. Finally, Appendix Figure A-2 plots the state-level data used to construct Panel A.

Panel B displays the inference procedure. The dark line is the difference between the actual and synthetic Colorado collections and should be viewed as the TABOR treatment effect. This line is nearly flat and deviates little from zero, which in the pre-TABOR period again demonstrates the success of the procedure at producing a credible counterfactual for Colorado, and in the post-TABOR period demonstrates the lack of any influence of TABOR on taxes. The lighter lines correspond to the placebo treatment effects; each one indicates the difference between collections in one of the 47 donor pool states and collections in the corresponding synthetically constructed state. (E.g. one of the light grey lines represents the difference between tax collections in Ohio and collections in a synthetic Ohio.) The thick mass of placebo effects concentrated around zero in the pre-TABOR period is encouraging as it suggests that the method generally produces synthetic tax collections capable of replicating actual tax collections.

In order to the reject the null hypothesis that TABOR had no effect, the line for Colorado would have to lie outside the mass of lighter lines in the post-period. As it lies well within the mass, we cannot reject the null. The dashed line displays the 90% percent confidence limit for the hypothesis that TABOR did not cause a decline in taxes (a one-sided test).[22]

[21] They are also presented in panel B of Table 1 for select years.

[22] The 95% percent confidence limit is similar, although a bit below, the 90% confidence interval. Available

Through 2000 we have the statistical power to detect a decline in excess of roughly \$230 with 90% percent confidence. (We interpret the size of this confidence limit below.)

If a synthetic state fails to fit well in the pre-treatment period, it is unlikely to provide useful information about the post-treatment period. In Panel B we omit such poor-fitting states by dropping any state with a pre-treatment mean squared predicted error (MSPE) in excess of five times Colorado's MSPE.[23] [24]

An alternative inference method involves calculating the ratio of the mean post-treatment MSPE to the mean pre-treatment MSPE (Abadie et al., 2010). Intuitively, a larger error—i.e. a larger difference between the actual and synthetic Colorado—in the post-treatment period than in the pre-treatment period is consistent with TABOR having caused a change in taxes. The Colorado effect is 'significant' if its MSPE ratio is large relative to the distribution of placebo state ratios—i.e. it lies in the upper end of the distribution. An advantage of this alternative inference approach is that it removes the need to exclude states with poor pre-treatment fits. Panel C of Figure 1 displays the distribution of ratios. The post-TABOR MSPE of Colorado is 5.7 times the pre-TABOR MSPE, placing Colorado in the lower end of placebo distribution. With a p-value of 0.60, we fail to reject the null.[25]

Given the magnitude of the TABOR rebates from 1997 through 2001—they ranged from \$35 to \$198 per person in these years—it may appear surprising that the synthetic collections replicate actual collections over this period. The puzzle is resolved by the fact that the synthetic Colorado was enacting large scale tax cuts during this period. Table 3 displays the cumulative reduction in per-capita tax liability due to changes in state government tax policy

---

from the authors upon request.

[23]Formally, $MSPE_i = (Z_1 - Z_0W)'(Z_1 - Z_0W)$. Abadie et al. (2010) omit states using cutoffs of twenty, five and two times the MSPE of the treated state. As the cutoff criteria becomes more stringent, the remaining pre-treatment fits are improved and provide greater assurance that post-treatment placebo effects are not artifacts of a poor fit. However, the quality of inference increases with the number of placebo states and there is therefore a cost to excluding states. We balance these conflicting interests by using a cutoff criteria of five times the MSPE of Colorado.

[24]The MPSE filter for the simultaneous results requires that the pre-treatment MSPE be less than five times the MPSE of Colorado for *both* the expenditure and tax outcomes. Thus, the inference bands for both outcomes are based on the identical set of placebo states.

[25]The p-value calculation is based on a two-sided test.

and, for Colorado, the TABOR rebates.[26] Although there are differences in timing, both states were reducing tax liability by a roughly similar amount over this period. A taxpayer in Colorado saw his 2001 liability fall by $250 relative to his 1995 liability. In the synthetic Colorado, the taxpayer experienced a nearly identical reduction of $240. While Colorado reduced its tax liability primarily through TABOR rebates, the synthetic state reduced its liability through legislated tax changes. In doing so, the two states maintained similar tax paths (Figure 1, Panel A). TABOR thus appears to have affected the administrative form of tax reduction, but not the fundamental level of taxes.[27]

Notably, the estimates on Panel B of Figure 1 are precise enough to rule out a decrease in taxes in Colorado of the magnitude reported on Table 3 in 2000 and 2001 with 90% confidence. The rebates and tax cuts for Colorado displayed in Table 3 are large in magnitude, equal to about $8\frac{1}{2}$ percent of total tax revenue in those years. The large magnitude of these tax reductions is typically viewed as a key piece of evidence in favor of the claim that TABOR suppressed revenue collections. It is therefore significant that the synthetic control procedure has sufficient statistical power to detect cuts of this size.

The tax results presented on Figure 1 are produced using the simultaneous weight method, with total government expenditures as the other outcome variable. The expenditure results are presented on Figure 2; they utilize the same synthetic Colorado as the previous tax results. The pre-TABOR expenditure fit is not nearly as tight as the pre-TABOR fit for taxes. Still, as can be seen in panel B, in the pre-TABOR period there is again a mass of placebo states around zero and Colorado is mostly within that mass. In the post-TABOR period, there is again an absence of any evidence that TABOR restricted the size of government. Through

[26] The table displays the effect of all revenue legislation taken by the state government on the *level* of per-capita tax liability in each year. A permanent tax cut of y dollars in 1996 is part of the effect in 1996 and all subsequent years. The TABOR rebates are simply entered as their value in the given year. Thus, the table displays the net reduction in tax liability in the state as the result of all state government tax policy changes from 1996 plus the effect of the rebates. The tax policy change data comes from the annual NASBO reports.

[27] Table 3 only includes information on changes in *state* government tax policy. It would be preferable to also include information on changes in tax policy at the local government. Unfortunately the required data is not available.

2000 we have the power to detect an a decline in expenditures due to TABOR in excess of about $430 per capita on average – equal to about 8% of expenditures over this period.

However, we detect no evidence of such a decline. In fact, the estimates are positive over essentially the entire post-TABOR period. Moreover, the null hypothesis that TABOR had no effect on expenditures can be rejected with 90% confidence in 2002 (two-sided test). Taken at face value this suggests that TABOR may have been seen by policy makers as an instruction (or permission) to grow government up to the legislated limits. If these limits were more permissive than what the legislature would have chosen in the absence of the policy, then TABOR would be expected to yield an increase in expenditure. However, the alternative inference (based on the MSPE pre and post treatment ratios) provides no support for the conclusion that TABOR boosted expenditures over the entire TABOR period—see Panel C. Colorado is in the middle of the distribution and the p-value of 0.30 suggests that TABOR had no effect on real expenditures per-capita in Colorado.[28]

Finally, our findings here stand in contrast to the past examinations of TABOR—discussed in section 1—which find that the policy reduced the size of government in Colorado. The likely explanation for the divergent findings is choice of control group. A strength of the synthetic control method is its rigorous, data driven choice of control group, as opposed to the more ad hoc strategies adopted in much of the past literature on tax and expenditure limitations.[29]

4.2 Floating and Priority Weight Results

We now turn to alternative possible approaches to estimating the effect of TABOR on taxes and expenditures using the synthetic cohort methodology. Figure 3 displays the results

[28]Capital expenditures are often quite lumpy from year to year. It is possible that the synthetic cohort procedure would have more success fitting current expenditures—total expenditures minus capital expenditures, interest on debt, assistance payments and insurance benefits—because of their lower annual volatility. Appendix Figure A-3 displays the results of using current expenditures and taxes as the outcome variables. There is a moderate improvement in the pre-TABOR fit for expenditures. Again there is no evidence that TABOR reduced taxes or expenditures.

[29]However, as presented in section 4.4.3, a carefully specified regression approach yields very similar findings to those we obtain from the synthetic control approach.

from the floating weight method – i.e. the standard synthetic cohort approach estimated separately for taxes and for expenditures. The results are quite similar to those produced by the simultaneous weight method and provide no evidence that TABOR reduced either taxes or expenditures.

Viewed in isolation, the floating weight result could be seen as a bit of a puzzle. Taxes and expenditures are closely linked through the budget identity and should broadly move together over time. However, the results provide no evidence that TABOR affected taxes, but provide at least some indication that it was associated with an increase in expenditures. One possible explanation is that the difference reflects the different counterfactual Colorados produced by estimating each outcome independently. The simultaneous weight procedure removes this possibility – a principal reason we prefer it over the floating weight approach.[30]

Appendix Figure A-4 displays the results of the priority weight method. The tax results (panel A) are based on the state weights from the standard procedure executed with expenditures as the outcome (displayed in panel B of Figure 3).[31] Correspondingly, the expenditure results (panel B) are based on the state weights from the standard procedure estimated with taxes as the outcome variable (displayed in panel A of Figure 3). These results provide no evidence that TABOR reduced taxes or expenditures.

Finally, Table 4 compares the three estimation approaches in terms of their ability to generate a synthetic Colorado which matches the actual Colorado in the pre-TABOR period (measured in terms of MSPE). As would be expected, estimation with the standard synthetic control approach separately for each outcome—i.e the floating weight approach—produces the lowest MSPE for both taxes and expenditures. Using the priority weight approach

[30]A different possible explanation for the divergence between the floating weight tax and expenditure results is that other forms of revenues such as fees, or the use of debt financing for capital expenditures, may have partially severed the connection between taxes and total expenditures. We return to this possibility below.

[31]The placebo states used to form the inference bands are handled analogously – i.e. they are based on the same synthetic states used to construct the expenditure inference bands. The maximum allowable pre-treatment MSPE filter for the inference bands is applied for expenditures only. The filter is not reapplied for taxes. Thus, the exact same set of placebo states appears in panel A of Figure A-4 as appears in panel B of Figure 3.

significantly degrades the ability of the procedure to match the non-prioritized outcome—e.g. for taxes the MSPE increases from 10,807 under the standard procedure to 206,072 under the priority weight method. The simultaneous weight approach provides a middle ground between the priority and floating weight approaches in terms of pre-period fit—e.g. an MSPE of 17,525 for taxes. We view the erosion in fit inherent in moving from the standard method to the simultaneous method as an acceptable trade off given the significant analytic advantage of holding the synthetic Colorado fixed across the two outcomes of interest.

4.3 Other Outcomes

Figure 4 contains the results for three additional revenue outcomes: fees, total government revenue and debt (all in per capita terms). In all cases, the state weights are taken from the simultaneous procedure run on taxes and expenditures (displayed in Figures 1 and 2). This approach is adopted in order to ensure that the counterfactual Colorado used to identify our preferred tax and expenditure results is also used to identify the effect of TABOR on other outcomes. We will refer to this procedure as the "fixed weight" approach because the weights are held fixed at the values generated by the simultaneous procedure.

The elevated black line in Panel A shows that fees in Colorado are somewhat higher than fees for the synthetic Colorado throughout the entire period, suggesting that Colorado has a preference for fees that the synthetic counterfactual consistently understates. Encouragingly, though, the trend is close to flat in the pre-period, suggesting that fees in the actual and synthetic Colorado evolved in a similar manner. There is a gradual upward drift starting in the second half of the 1990s, but the accompanying histogram in the right of Panel A—which gives us some idea of significance over the entire period—fails to provide evidence of an effect. Furthermore, we acknowledge that the inference in the left of Panel A is highly problematic due to the divergence in levels between actual and synthetic fees in the pre-period.[32]

The results for total government revenues—which include revenues such as fees and grants

[32] One approach would be to simply remove the inference bands from Figure 4. While we are sympathetic to this view, some readers may wish to view the bands and we have therefore retained them.

from the federal government, as well as taxes—are displayed in Panel B. Total government revenues in Colorado evolved similarly to those in the synthetic Colorado from 1977 through the early 2000s when revenues in Colorado begin to outstrip those of the synthetic state. However, for the entire post-TABOR period the MSPE ratio fails to provide support for a TABOR effect (see the right panel).

Using state and local government debt per-capita as an outcome produces a poor and inconsistent fit in the pre-period so we cannot say much about it with any confidence (panel C). It is interesting that in the pre-TABOR period, debt per-capita in Colorado is smaller than in the synthetic Colorado, but the negative gap diminishes and eventual becomes positive by the early 2000s. The upward trend, though, appears to have started before TABOR suggesting it is not causally connected to the policy.

Even if TABOR does not alter the overall level of spending, it may have altered the composition of spending. Many observers have argued that TABOR shifted the composition of state spending away from discretionary portions of the budget, such as public health, and toward non-discretionary areas, such as corrections (which are largely driven the prison population, not the annual budgeting process). Figure 5 presents two spending categories: K-12 Education and Health and Hospitals. No evidence is found that TABOR influenced these outcomes and extensive examination of additional budget categories (unreported) similarly fails to find evidence of a TABOR effect.

4.4 Robustness and Assessment of Role of Non-TABOR Fiscal Rules

This section serves two purposes. First, a key advantage of the synthetic control method is its data-driven choice of a counterfactual. Still, it remains an unverifiable assumption that the synthetic Colorado is an appropriate counterfactual for Colorado in the period after TABOR. We therefore execute numerous robustness checks. Second, to properly interpret the TABOR treatment effects estimated above, it is important to assess the role of possibly binding fiscal rules in the control states. This section provides such an assessment. There are three subsections: The first presents specifications using our primary state-level synthetic control approach; the second presents synthetic control estimates based on county-level data; the third presents results from a regression-based dynamic difference-indifference regression approach.

4.4.1 State-level Synthetic Control Estimates

We conduct three types of robustness checks using our primary state-level synthetic control approach. First, one might worry that changing the set of fiscal predictors in X might generate different conclusions. We therefore construct an alternative synthetic Colorado using an extremely parsimonious X vector which contains only indicator variables for Census division. The synthetic Colorado is therefore composed of equally weighted Mountain Division states (excluding Colorado).[33] These states share many of the geographic, economic, demographic and political characteristics of Colorado. Panel A of Figure 6 shows the results for taxes. While the pre-TABOR fit isn't particularly good, as Colorado has persistently higher taxes than the average Mountain state, the gap between the synthetic and real Colorado's is fairly constant throughout the seven years leading up to TABOR and in most of the post-TABOR period. Thus, TABOR appears to have had no effect on taxes in Colorado. Panel B similarly displays no evidence of an effect on expenditures.

Second, one possible objection to our approach is its failure to account for potential spillover effects of TABOR into bordering states.[34] Indeed, 76% of the weight in the primary synthetic Colorado is assigned to states that border Colorado. If these states exhibit a behavioral response to TABOR it may bias our estimates. For instance, if policy makers in neighboring states fear a TABOR-like policy may be passed by their voters, they may reduce taxes in an effort to signal the lack of need for such a policy. To address this concern, we run the simultaneous method after dropping all of the states bordering Colorado from the donor pool. As is visible in panel C of Figure 6, there is a distinct downward drift in the tax collections of Colorado compared to the synthetic state. The descent begins around 1985

[33]Three of the six states in the primary synthetic Colorado are in the Mountain division—see Table 2.
[34]Baicker (2005) and Case et al. (1993) provide evidence of policy spillovers between neighboring states.

and ends around 1996. (The downward tilt is made clear in panel A of Appendix Figure A-5 which adds a trend line to visually reveal the almost perfectly linear descent of the treatment effect from the mid-1980s through the mid-1990s.) In order to credibly conclude that TABOR caused the change, the trend break would need to occur at the time of TABOR went into effect or afterwards. Thus, we cannot conclude that TABOR *caused* the decline in taxes. Panel D indicates that the synthetic state does a very poor job of matching expenditures in the pre-TABOR period. While this severely limits the ability to draw conclusions, there is no evidence that TABOR pushed down expenditures.

Third, it is important to quantify the extent to which our TABOR treatment effect reflects binding (or stringent) fiscal rules in the control states. We are estimating the effect of TABOR against a counterfactual of not having a TABOR-like policy. If a synthetic state passes an effective non-TABOR fiscal rule in the post-TABOR period this will reduce synthetic tax collections and reduce the magnitude of the TABOR treatment effect. Thus, one possible explanation for the lack of a negative TABOR treatment effect is the enactment of effective fiscal rules in the control states. The enactment of fiscal rules was quite common in the sample period, as 31 states enacted either a TEL or a legislative supermajority requirement in 1977 or later.

To examine the role of non-TABOR fiscal rules in our primary TABOR estimates, we estimate the simultaneous method excluding from the sample states which either passed a stringent general TEL or imposed a legislative supermajority requirement for tax increases in 1987 or later. (See appendix section 6.3 for information on how these stringent fiscal rule variables are defined.) This restriction leaves only 30 states in the donor pool.[35] Likely due to this smaller donor pool, the pre-treatment fit of the procedure is somewhat reduced for taxes (panel E). That said, there is no evidence of a decrease in taxes in the post-TABOR period

[35]The sample selection criteria eliminates three states which contribute to the synthetic Colorado for the results in Figures 1 and 2: Arizona enacted an expenditure TEL in 1979 and a legislative supermajority requirement in 1992; Nevada enacted an expenditure TEL in 1979 and a legislative supermajority requirement in 1996 and Utah enacted a expenditure TEL in 1989. As discussed above, Colorado enacted an expenditure TEL in both 1977 and 1991.

and the 90% confidence limit is actually a bit smaller through 2005 than in the primary estimates. Similarly, there is no evidence that TABOR depressed expenditures (panel F). Thus, these estimates suggest that the failure to find evidence that TABOR decreased taxes and expenditures is not primarily explained by binding non-TABOR fiscal rules in the control states.

Finally, in the Appendix we consider two additional types of robustness checks. The first concerns how we measure our fiscal outcomes. In all cases we have measured budget outcomes in per capita terms. An alternative possibility is to use the log of the budget outcomes. Normalizing by personal income is another possible alternative. We explore these alternatives in Appendix section 6.4. The estimates using logs produce conclusions very similar to those produced using per capita outcomes. The estimates using budget outcomes as a fraction of personal income must be interpreted with care due to the presence of pronounced downward trends in the estimates in the pre-TABOR period. That said, these estimates fail to produce evidence that TABOR decreased taxes or spending. The second set of Appendix robustness checks examine the sensitivity of the estimates to variation in the set of X fiscal predictors – see Appendix section 6.2. The TABOR treatment effect estimates are shown to be relatively insensitive to the precise set of variables chosen for inclusion in X.

4.4.2 County-Level Synthetic Control Estimates

We have argued that an analysis of TABOR requires a joint evaluation of all layers of sub-national government – hence our use of state and local government tax and expenditures aggregated to the state level. That said, a possible drawback of our preferred approach is that around 75% of the control group is comprised of only 3 states. A single unobserved shock in one of these states could conceivably bias our estimated effects of TABOR. To address this possible concern, we perform an analysis of local government budget outcomes based on an extremely geographically diffuse control group.[36]

[36] A diffuse control group is also useful in assessing the influence of fiscal rules in the control states on our primary synthetic control estimates. Even if non-TABOR fiscal rules do not bind on average, a single binding fiscal rule in a state which contributes to the synthetic Colorado could heavily influence our state-level results.

Our approach here is similar in spirit to the work of Kline and Moretti (2014) who evaluate the economic effects of the Tennessee Valley Authority (TVA). Like TABOR, the TVA involves a single treatment unit – the Tennessee Valley. Unlike the primary analysis in this paper, though, the analysis in Kline and Moretti (2014) is conducted at the county level by matching TVA counties to observationally similar non-TVA counties.

We adopt a similar county-level approach using the synthetic control method. We employ data on all local government taxes and expenditures aggregated to the county level by the Census Bureau (only available in years ending in 2 and 7 when a census of all governments is taken) and a set of county-level fiscal predictors X equivalent to the state-level predictors on Table 1.[37]

For each of the 63 counties in Colorado we generate a synthetic counterpart from the nearly 3,000 counties in other states.[38] We do so using our simultaneous weight version of the synthetic control method with taxes and expenditures per capita as the outcome variables. In order to perform inference, we execute the same county-level procedure on non-Colorado counties. Similar to the approach developed in Acemoglu et al. (2016), we take population weighted averages of the county-level treatment effects by state to form both the TABOR treatment effect (from the Colorado counties) and the inference bands (from the non-Colorado states). As the fiscal outcome variables are in per-capita terms, the population weighted average yields the per-capita treatment effect for the state as a whole.[39] Given the

Such a scenario would almost certainly have a much smaller effect on the TABOR treatment effect when a very diffuse control group is used.

[37]The "political variable" set on Table 1 is omitted because there is no readily available equivalent at the county level.

[38]In 2001, Broomfield County, CO was created from of pieces of Adams, Boulder, Jefferson and Weld counties. It is a moderately large county with a population of a around 40,000 in 2001 – equal to around the 75th percentile of the Colorado county population distribution in that year. Broomfield did not exist in the pre-TABOR period and it therefore cannot be used in the estimation procedure (as the synthetic control procedure only uses data from the pre-TABOR period). We therefore allocate the taxes and expenditures of Broomfield to the counties from which it was created in proportion to the population contributed to the new county. We drop counties with border changes in other states from the sample.

[39]For the population weights, we use population as measured in 1992 – the year before TABOR was implemented. In addition to yielding the per-capita treatment effect for the state as a whole, the population weighted averaging also addresses the often poor pre-TABOR fit of the synthetic control procedure for very small counties. About one-third of Colorado counties had a population of less than 5,000 in 1992.

large size of the donor pool, this is an extremely computationally intensive procedure.

Figure 7 displays the effective weights assigned by the procedure to all non-Colorado counties. As can be seen, the control group is very geographically diverse. As would be expected, Mountain division counties receive substantial weight. However, counties in the East North Central, Middle Atlantic, New England divisions and counties in portions of the Pacific and West North Central also receive significant weight.

Figure 8, panel A displays the results for the tax outcome. The treatment effect is extremely flat throughout the pre and post TABOR periods, providing no evidence that TABOR decreased local government tax collections. This finding is quite consistent with the prevalence of successful TABOR override votes discussed in section 1.2. Panel B displays the results for the expenditure outcome. Although the treatment effect bounces around over time (decreasing over the decade from 1987 to 1997 and them increasing over the following decade), there is no clear overall trend and no evidence that TABOR decreased local government expenditures.

4.4.3 Regression-Based Dynamic Difference-in-Difference Estimates

We have argued that the synthetic control approach has significant advantages relative to more traditional regression-based approaches for evaluating the effects of TABOR. Nonetheless, regression-based estimates have clear value-added in this setting for two reasons. First, they demonstrate the robustness of our conclusions to an alternative, more familiar, estimation technique. Second, a regression based approach, unlike the synthetic control method, can directly control for time-varying observables in the post-treatment period; in particular, it can control for the effect of fiscal rule implementation in the control states.

We employ a dynamic difference-in-difference (DD) research design:

$$y_{it} = \sum_{g=-15}^{20} \beta_g TABOR_{g,it} + \eta_i + \zeta_{jt} + \epsilon_{it} \tag{8}$$

where y_{it} is the outcome variable (total state and local government taxes or expenditures per capita) for state i in fiscal year t, η_i is a vector of state fixed-effects, ζ_{jt} is a vector of Census division j - fiscal year t fixed-effects, and $TABOR_{g,it}$ is a dummy variable equaling one if state i at time t had enacted TABOR g years ago, with $g = 1$ denoting the year of passage. The year prior to passage ($g = 0$) is the omitted category.

The state fixed-effects, η_i, control for time-invariant state characteristics. The division-year fixed-effects, ζ_{jt}, control for shocks common to states at the region-year level such as shifts in economic conditions. The β vector is the parameter of interest. It traces out the dynamic treatment effect of TABOR.

The identifying assumption of the model is that, absent TABOR passage, Colorado would have had tax collections similar to the control states, conditional on the covariates. Underlying trends in the outcome variable are perhaps the most likely violation of this assumption. The pre-TABOR portion of the β vector provides a check against this possible violation. If TABOR is unassociated with underlying trends, there should be no trend in the β vector in pre-TABOR period.

Confidence intervals for state-level DD estimates are typically based on standard errors clustered by state. However, such confidence intervals are possibly, but not necessarily, biased when there is only a single treated unit (Conley and Taber, 2011). To account for this possibility, we adopt a conservative approach and construct a Conley and Taber (2011) lower-bound confidence limit (one-sided test).[40]

Panel A of Figure 9 presents the results of estimating equation (8) with year fixed effects (as opposed to the year-division effects). Several of the post-TABOR years display a negative treatment effect. However, the pre-TABOR treatment vector reveals a distinct downward trend in state and local government tax collections in Colorado prior to TABOR's passage which raises serious doubts about the validity of the model's identifying assumption. A more simplistic approach which uses a single post-TABOR indicator variable and standard errors

[40]Conley and Taber (2011) confidence intervals, like the synthetic control confidence intervals, are based on a form of exact, or permutation, inference (Rosenbaum, 2002); they are appropriate for a setting in which only a single unit has been treated. Confidence intervals for the results in Figure 9 based on standard errors clustered by state are substantially narrower.

clustered by state produces a TABOR treatment effect of negative \$150 per capita which is significant at the 1 percent level (available from the author's upon request). The results in Panel A, though, cast significant doubt on a causal interpretation of this estimate.

Panel B presents results with year-Census region fixed-effects. The troubling pre-TABOR downward trend remains. Panel C employs Census division-year fixed-effects (i.e. exactly equation (8)). This specification produces an essentially flat pre-trend from the early 1980s through the passage of TABOR. Viewed jointly, panels A-C suggest that, conditional on state and year fixed effects, Mountain division states (including Colorado) had elevated but declining tax collections in the years prior to TABOR. This division-level trend is inappropriately ascribed to the TABOR treatment effect vector in the absence of the Census division-year controls. We therefore prefer specifications with division-year controls.

Panel C produces no evidence that TABOR depressed tax collections and the treatment effect in terms of both magnitude and contour is remarkably similar to that produced by the synthetic cohort procedure in Figure 1. Moreover, despite the conservative Conley and Taber (2011) inference procedure, the estimates are precise enough to rule out a large negative effect. For instance, from 1993 to 2000 the negative effect of TABOR on tax collections can be bounded at around \$145 per capita on average with 90% percent confidence – equal to only about $5\frac{1}{2}$% of average per capita tax collections over this period. The confidence limit is particularly tight from 1998 to 2004 when the negative TABOR effect can be bounded at about a 4% decline in average per capita taxes.

Panels D and E control for the effects of non-TABOR fiscal rules which are deemed to be strict – an important robustness check given the inability of the synthetic control estimates to control for the introduction of such rules in the post-TABOR period. In panel D, we control with indicator variables for the presence of four types of stringent fiscal rules: general revenue limits, statutory expenditure limits, constitutional expenditure limits, and tax legislative supermajority requirements. In panel F we control for these same four variables in an extremely flexible manner by allowing their effect to vary by time relative to implementation – i.e. we control for four terms: $\sum_{h=-10}^{10} \pi_h Rule_{h,it}$ where $Rule_{h,it}$ denotes a stringent fiscal rule in state i at time t implemented h years ago.[41] The inclusion of the controls for stringent fiscal rules has almost no effect on the TABOR treatment effect point estimates. Results are similar when controlling for the stringency of state-level fiscal rules using the ordinal rankings developed by Resnick (2002) (available from the authors upon request). The near invariance of the estimates to very flexible fiscal rule controls is relatively direct evidence that the TABOR treatment effect is not explained by binding fiscal rules in the control states.

Finally, panel F displays the results for expenditures controlling for division-year effects and the flexible fiscal rule vectors (i.e. the same specification as in panel E). Similar to the results elsewhere in the paper, the expenditure results are more variable from year to year than are the tax results. That said, there is no evidence that TABOR reduced expenditures. The contour of the treatment effect is again quite similar to that produced by the synthetic control approach – see Figure 2.

5 Conclusion

We find no evidence that TABOR influenced budget outcomes. The failure of TABOR to constrain the growth of the public sector likely reflects three factors. First, the policy appears to have been broadly consistent with the preferences of policy makers for at least a portion of the post-TABOR period. In the period between enactment in 1992 and 2001, Colorado issued large TABOR rebates to it citizens. However, the citizens of the synthetic Colorado were similarly receiving large tax reductions in the form of legislated tax cuts. TABOR dictated the unusual form the tax reductions took, but apparently not the fundamental level of taxes or spending. Thus, TABOR appears to be not a cause of the size of government, but merely a ratification of the state's preferences regarding the size of its public sector.

Second, policy makers likely engaged in some subversion of the limits, particularly in the

[41] $Rule_{-10,it}$ equals one for all years in which it is 10 *or more* years till passage and $Rule_{10,it}$ equals one for all years in which it is 10 *or more* years after passage.

later years of the policy. While it is not possible to identify all instances of such subversion or quantify their magnitude, there is at least one significant example: Following Amendment 23 and its mandate of increased funding for K-12 education, higher education was slated to see a diminishing share of the state budget. The legislature at least partially avoided these cuts through a somewhat complex set of changes involving replacing direct appropriations for higher education with stipends issued directly to students (to be used for tuition) and fee-for-service contracts between the state and public higher education institutions. These changes allowed the state's colleges and universities to be reclassified as enterprises and hence exempt from TABOR.[42] Significantly, increases in tuition and other forms of university revenue were no longer included under the global revenue growth cap. Given the rapid rise in tuition during this period, removing these revenues from the cap eased pressure on both higher education and the state budget as a whole.

Third, voters appear to have repeatedly suffered buyers remorse and relaxed the constraints they had placed on policy makers. At the local level, a stream of overrides were passed throughout the TABOR period as individual communities found the limit too restrictive. At the state level, voters can directly take on the role of policy maker through the referendum process. Twice these policy makers chose to undermine the intent of their predecessors who passed TABOR. Amendment 23 increased permissible revenue growth and mandated increased K-12 education spending.[43] TABOR was set to bind rather tightly in fiscal 2006 had voters not temporarily removed the revenue growth limit by passing Referendum C. At no point did voters revoke TABOR. (For instance, Referendum C only temporarily removed the growth cap and retained other elements of the policy.) They did, however, apparently loosen it sufficiently to allow government to maintain the path it would have taken in the absence of the policy.

It is not possible to place precise weights on the importance of these three factors. Our

[42]See Hill (2004); Eslinger (2005); Symanski (2010).

[43]However, Amendment 23 influenced tax collections only in fiscal 2001, as the subsequent economic downturn drove revenues well below the level required for rebates to be issued in fiscal 2002 and beyond.

reading of the qualitative evidence, though, suggests that the first and third factors were likely more important than the second. TABOR was constructed quite stringently and left only limited scope for the type of evasion that occurred in the area of higher education. Regardless of the relative importance of the three factors, our results are very consistent with the institutional irrelevance theory of fiscal rules and provide no support to the claim that such rules constrain budget outcomes. We again caution, though, that as TABOR has been applied in only a single state, a degree of unavoidable uncertainty exists with regard to the external validity of our conclusions.

Finally, the conclusion that TABOR did not affect taxes or expenditures is insufficient to conclude that such policies have no welfare effects. On one hand, these types of limitations may increase the odds that the provision of public goods deviates from that desired by the public. For example, in several instances when TABOR was set to bind, it was circumvented by voters. It is possible that in the future similar attempts by voters to circumvent such policies may fail in a manner which reduces welfare. It is also possible that TABOR effectively increased the deadweight loss generated by the state tax system. In particular, the revenue used to fund the TABOR rebates was raised through distortionary taxation. In contrast, the synthetic Colorado never raised this revenue (i.e. it engaged in legislative tax reductions instead). Thus, despite having similar tax collections net of the rebates, the synthetic Colorado's tax system likely generated less deadweight loss. On the other hand, even if such policies never affect government outlays or revenues, they may increase welfare by reducing the uncertainty faced by households and businesses over the size of the public goods bundle and tax burdens (Baker et al., 2016).

References

Abadie, Alberto, Diamond, Alexis, and Hainmueller, Jens, 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program." *Journal of the American Statistical Association* 105(490): 493–505.

Abadie, Alberto, Diamond, Alexis, and Hainmueller, Jens, 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2): 495–510.

Abadie, Alberto and Gardeazabal, Javier, 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1): 113–132.

Acemoglu, Daron, Johnson, Simon, Kermani, Amir, Kwak, James, and Mitton, Todd, 2016. "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics* 121(2): 368 – 391.

ACIR, 1995. "Tax and Expenditure Limits on Local Governments." Tech. rep., Advisory Commission on Intergovernmental Relations.

Anderson, Nathan, 2006. "Property Tax Limitations: An Interpretative Review." *National Tax Journal* .

Ando, Michihito, 2015. "Dreams of urbanization: Quantitative case studies on the local impacts of nuclear power facilities using the synthetic control method." *Journal of Urban Economics* 85: 68–85.

Athey, Susan and Imbens, Guido W., 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2): 3–32.

Baicker, Katherine, 2005. "The Spillover Effects of State Spending." *Journal of Public Economics* 89: 529–544.

Baker, Scott R., Bloom, Nicholas, and Davis, Steven J., 2016. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics (forthcoming)* .

Baron, Janis, Dunn, Tom, and et. al., 2003. "House Joint Resolution 03-1033 Study: TABOR, Amendment 23, the Gallagher Amendment, and Other Fiscal Issues." Tech. rep., Legislative Council of the Colorado General Assembly.

Blom-Hansen, Jens, Bkgaard, Martin, and Serritzlew, Sren, 2014. "Tax Limitations and Revenue Shifting Strategies in Local Government." *Public Budgeting & Finance* 34(1).

Bohn, Sarah, Lofstrom, Magnus, and Raphael, Steven, 2014. "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *Review of Economics and Statistics* 96: 258–269.

Bridges, Rutt, 2004. "Yes: Colorado Needs to Get Its Budget In Order." Denver Post.

Brooks, Leah, Halberstam, Yosh, and Phillips, Justin, 2012. "Spending within Limits: Evidence from Municipal Fiscal Restraints." University of Toronto Working Paper.

Card, David, 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 44: 245–257.

Card, David and Krueger, Alan, 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84: 772–793.

Case, Anne, Rosen, Harvey, and Hines, James, 1993. "Budget spillovers and fiscal policy interdependence: Evidence from the states." *Journal of Public Economics* 52: 285–307.

Cavallo, Eduardo, Galiani, Sebastian, Noy, Ilan, and Pantano, Juan, 2013. "Natural Disasters and Economic Growth." *Review of Economics and Statistics* 95: 1549–1561.

Colorado Fiscal Policy Institute, 2004. "TABOR - A Brief Outline." Tech. rep., Colorado Center on Law and Policy.

Colorado Municipal League, 2012. "Colorado Municipal Facts." Tech. rep.

Conley, Timothy G. and Taber, Christopher R., 2011. "Inference with Difference in Differences with a Small Number of Policy Changes." *The Review of Economics and Statistics* 93(1): 113–125.

Cutler, David, Elmendorf, Douglas, and Zeckhauser, Richard, 1999. "Restraining the Leviathan: Property Tax Limitations in Massachusetts." *Journal of Public Economics* 71(3): 313–334.

Dustmann, Christian, Schnberg, Uta, and Stuhler, Jan, 2017. "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment*." *The Quarterly Journal of Economics* 132(1): 435–483.

Dye, Richard, McGuire, Therese, and McMillen, Daniel, 2005. "Evidence on the Short and Long Run Effects of Tax Limitations on Taxes and Spending." *National Tax Journal* 58(2).

Eslinger, Cathy, 2005. "The College Opportunity Fund: A New Approach to Higher Education Funding." Colorado Legislative Council Staff Issue Brief.

Frates, Chris, 2005. "Fiscal Folly?" *State Legislatures* pages 20–23.

Greenwood, Daphne and Brown, Tom, 2000. "An Overview of Colorado's State and Local Tax Structure." Tech. rep., Center for Colorado Policy Studies.

Hedges, Carol, 2003. "Ten Years of TABOR." Tech. rep., The Bell Policy Center.

Hill, Joanne, 2004. "University of Colorado Enterprise Designation." Memorandum from Office of State Auditor to Members of the Legislative Audit Committee.

Hinrichs, Peter, 2012. "The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities." *Review of Economics and Statistics* .

James, Franklin and Wallis, Allan, 2004. "Tax and Spending Limits in Colorado." *Public Budgeting and Finance* pages 16–33.

Kline, Patrick and Moretti, Enrico, 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *." *The Quarterly Journal of Economics* 129(1): 275–331.

Kousser, Thad, McCubbins, Mathew, and Moule, Ellen, 2008a. "For Whom the TEL Tolls: Can State Tax and Expenditure Limits Effectively Reduce Spending?" *State Politics & Policy Quarterly* 8: 331–361.

Kousser, Thad, McCubbins, Mathew, and Rozga, Kaj, 2008b. "When Does the Ballot Box Limit the Budget? Politics and Spending Limits in California, Colorado, Utah and Washington." In Elizabeth Garret, Elizabeth Graddy, and Howell Jackson, (Eds.) "Fiscal Challenges: An Interdisciplinary Approach to Budget Policy," New York, NY: Cambridge University.

Lav, Iris, 2009. "Maine and Washington Reject TABOR." Tech. rep., Center for Budget and Policy Priorities.

Lav, Iris and Williams, Erica, 2010. "A FORMULA FOR DECLINE: Lessons from Colorado for States Considering TABOR." Tech. rep., Center for Budget and Policy Priorities.

Legislative Council Staff, 2001. "Amendment 23: A Breif Overview." Tech. rep., Legislative Council of the Colorado General Assembly.

Martell, Christine R. and Teske, Paul, 2007. "Fiscal Management Implications of the TABOR Bind." *Public Administration Review* 67: 673–687.

McGuire, Therese and Rueben, Kim, 2006. "The Colorado Revenue Limit: The Economics Effects of TABOR." *Economic Policy Institute Briefing Paper* 18: 233–246.

New, Michael, 2010a. "The Return of TABOR." Tech. rep., Cato Institute.

New, Michael and Slivinski, Stephen, 2005. "TABOR Time." Tech. rep., Cato Institute.

New, Michael J., 2010b. "U.S State Tax and Expenditure Limitations: A Comparative Political Analysis." *State Politics & Policy Quarterly* 10(1): 25–50.

New York Times, 2015. "The Glitch in Colorado's Weed Experiment." New York Times.

Peri, Giovanni and Yasenov, Vasil, 2015. "The Labor Market Effects of a Refugee Wave: Applying the Synthetic Control Method to the Mariel Boatlift." Working Paper 21801, National Bureau of Economic Research.

Poterba, James, 1997. "Do Budget Rules Work?" In Alan Auerbach, (Ed.) "Fiscal Policy: Lessons from Economic Research," Cambridge, MA: MIT Press.

Poterba, James and Rueben, Kim, 1995. "The Effect of Property-Tax limits on Wages and Employment in the Local Public Sector." *American Economic Review* 85(2): 384–389.

Poterba, James and Rueben, Kim, 1999. “Fiscal Rules and State Borrowing Costs: Evidence from California and Other States.” Public Policy Institute of California.

Poterba, James and von Hagen, Jurgen, 1999. “Introduction.” In James amd Jurgen von Hagen Poterba, (Ed.) “Fiscal Institutions and Fiscal Performance,” Chicago, IL: University of Chicago Press.

Poulson, Barry, 2005a. “Tax and Expenditure Limits: Experiments in Direct Democracy.” Tech. rep., Americans for Prosperity.

Poulson, Barry, 2005b. “Understanding the Attack on TABOR.” Tech. rep., Americans for Prosperity.

Resnick, Phyllis, 2002. “Fiscal Cap Style TELs in the States: An Inventory and Evaluation.” Tech. rep., Center for Tax Policy.

Rosenbaum, Paul R., 2002. “Covariance Adjustment in Randomized Experiments and Observational Studies.” *Statistical Science* 17(3): 286–327.

Rueben, Kim, 1997. “The Effect of Tax and Expenditure Limits on State and Local Governments, Chapter 1.” Massachusetts Institute of Technology.

Skidmore, Mark, 1999. “Tax and expenditure limitations and the fiscal relationships between state and local governments.” *Public Choice* 99(1): 77–102.

Symanski, Sally, 2010. “Higher Education TABOR Enterprise Status.” Memorandum from Office of State Auditor to Members of the Legislative Audit Committee.

Waisanen, Bert, 2010. “State and Tax Expenditure Limits - 2010.” Tech. rep., National Conference of State Legislatures.

Wall Street Journal, 2002. “State of Prosperity (or Not).” Wall Street Journal.

Wall Street Journal, 2004. “A Tax and Spend Lesson.” Wall Street Journal.

Watkins, Kate, 2003. “House Joint Resolution 03-1033 Study: TABOR, Amendment 23, the Gallagher Amendment, and Other Fiscal Issues.” Tech. rep., Legislative Council of the Colorado General Assembly.

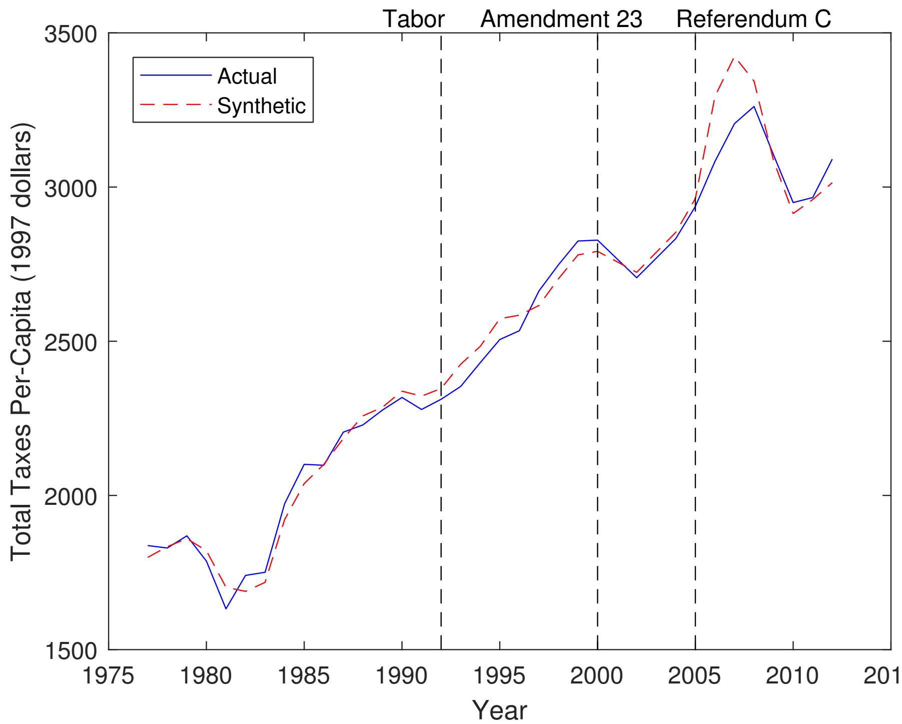
Will, George, 2005. “Taxes and Colorado’s Apostate.” Washington Post.

Will, George, 2011. “In Colorado, Colliding Views on the Proper Form of Government.” Washington Post.

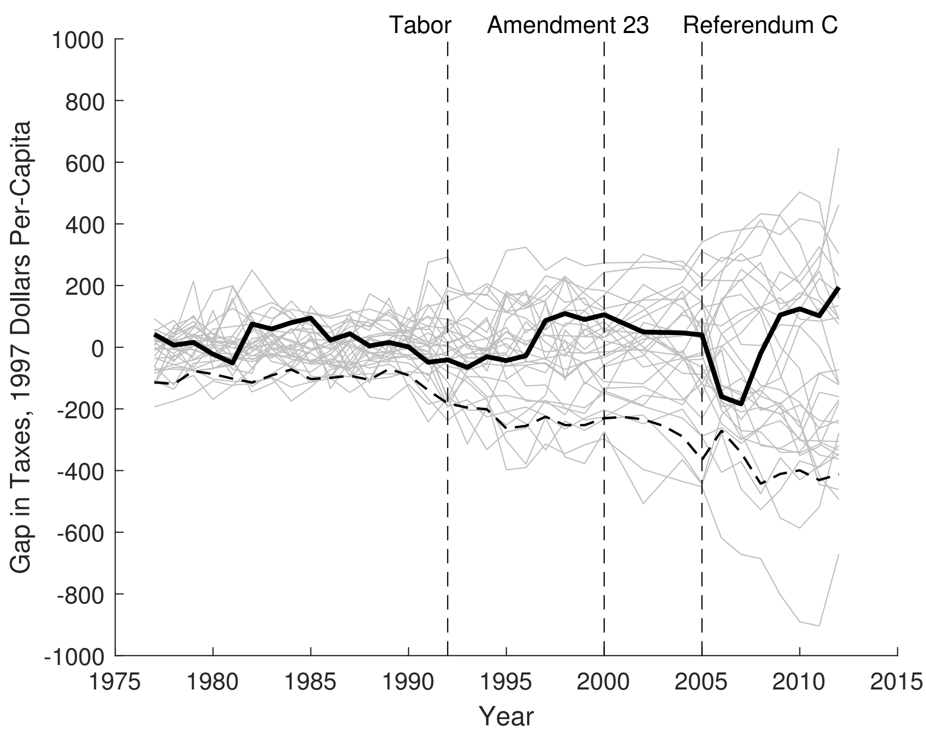
Zhang, Pengju, 2017. “The unintended impact of tax and expenditure limitations on the use of special districts: the politics of circumvention.” *Economics of Governance* .

Figure 1: Taxes Per Capita

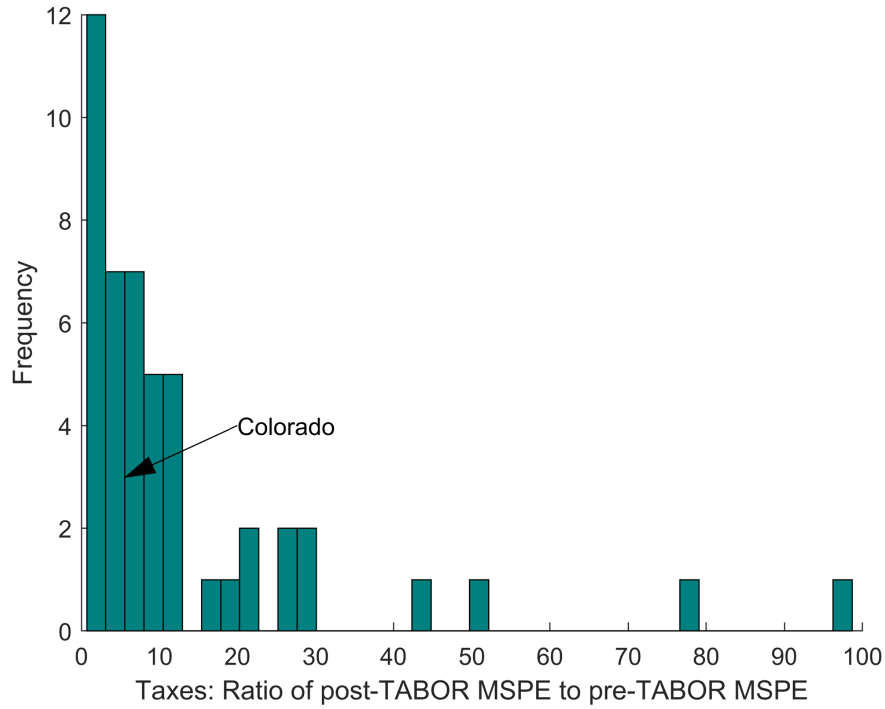
Panel A



Panel B

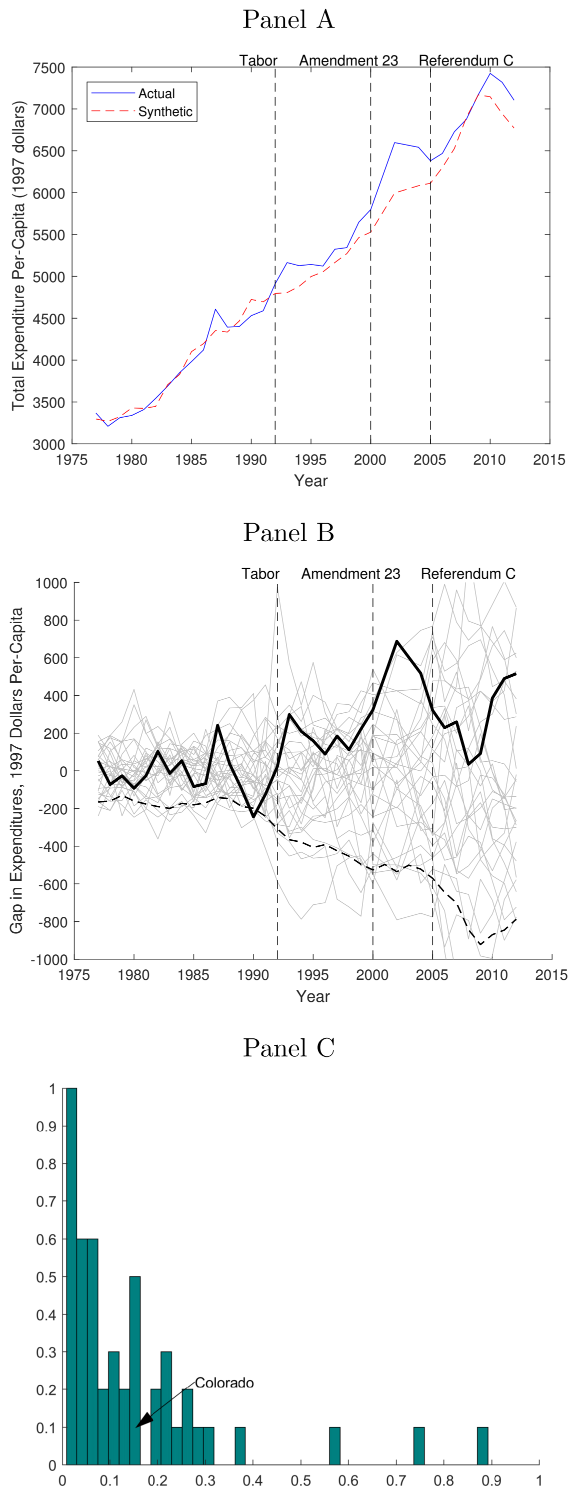


Panel C



Note. Simultaneous weight method; total government expenditures per capita used as the other outcome variable (displayed in Figure 2). Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in panel B. All control states are included in Panel C. The dashed line in panel B displays the lower bound 90% percent confidence limit (one-sided test).

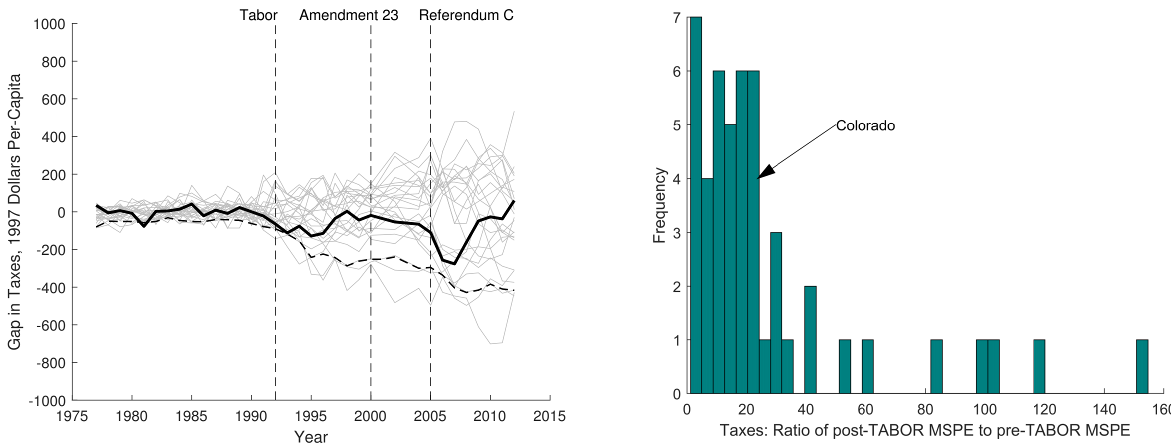
Figure 2: Total Expenditures Per Capita



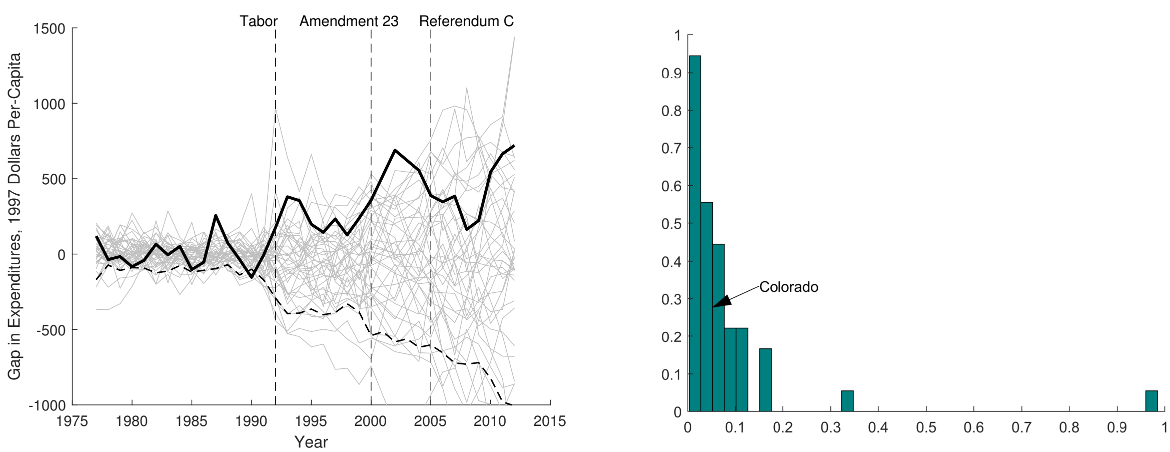
Note. Simultaneous weight method; taxes per capita used as the other outcome variable (displayed in Figure 1). Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in Panel B. All control states are included in Panel C. The dashed line in panel B displays the lower bound 90% percent confidence limit (one-sided test).

Figure 3: Floating Weight Method

Panel A: Taxes Per Capita



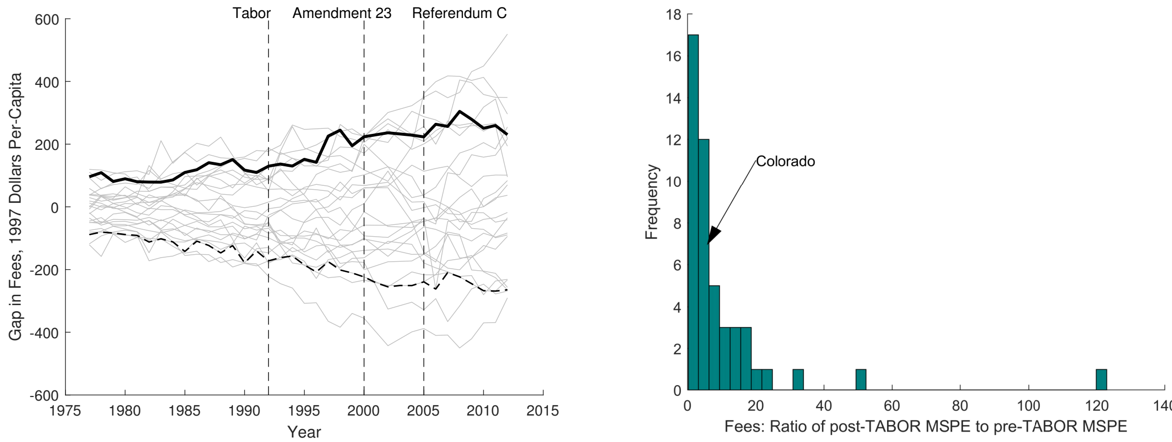
Panel B: Total Government Expenditures Per Capita



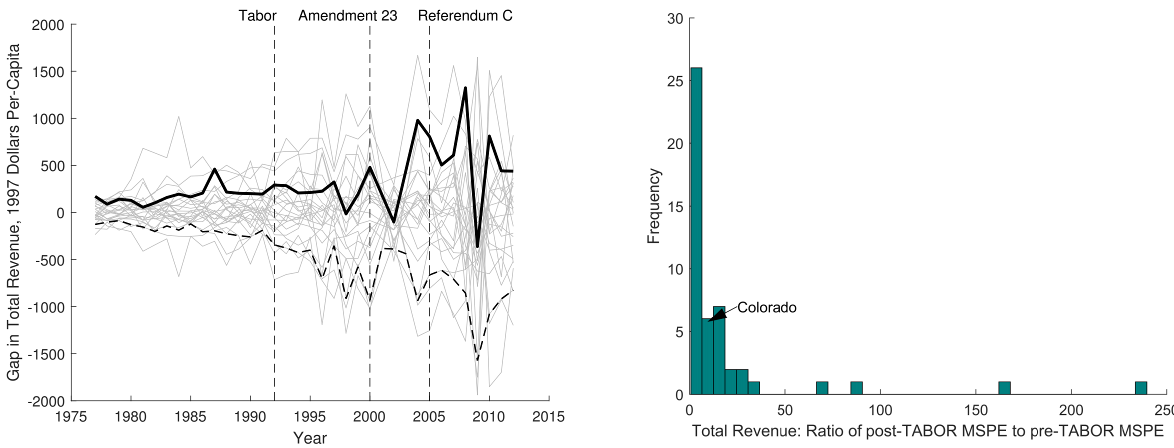
Note. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in the left-hand figures. All control states are included in the right-hand side figures. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Figure 4: Fees, Total Revenue and Debt

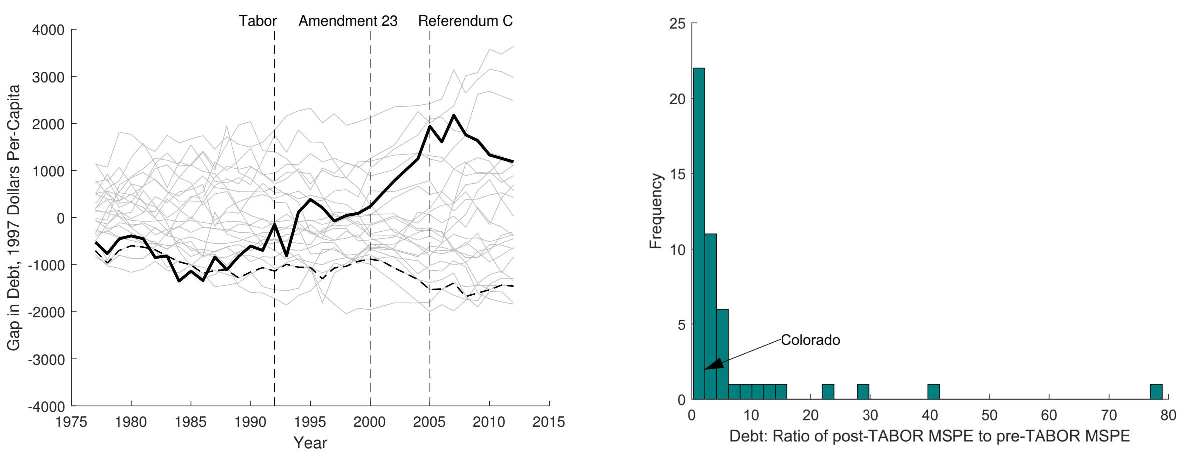
Panel A: Fees Per Capita



Panel B: Total Government Revenues Per Capita



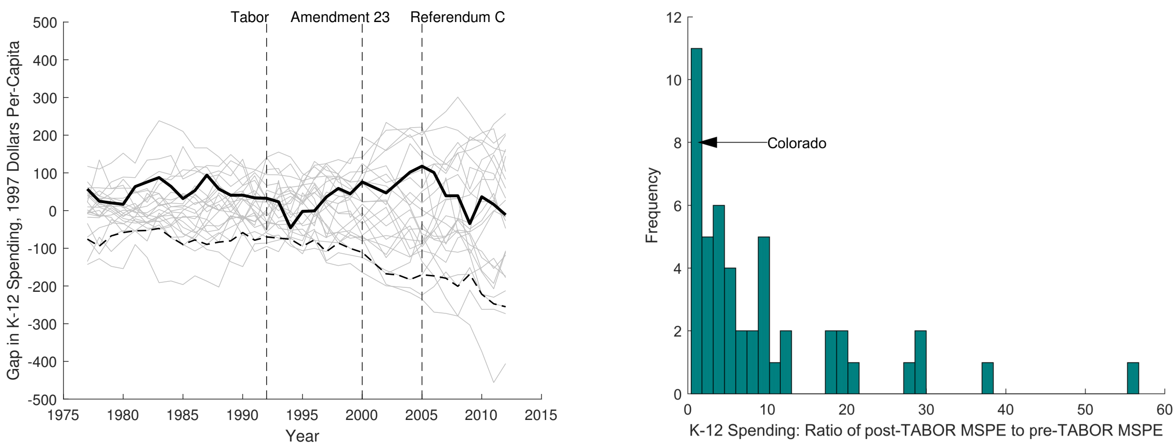
Panel C: Debt Per Capita



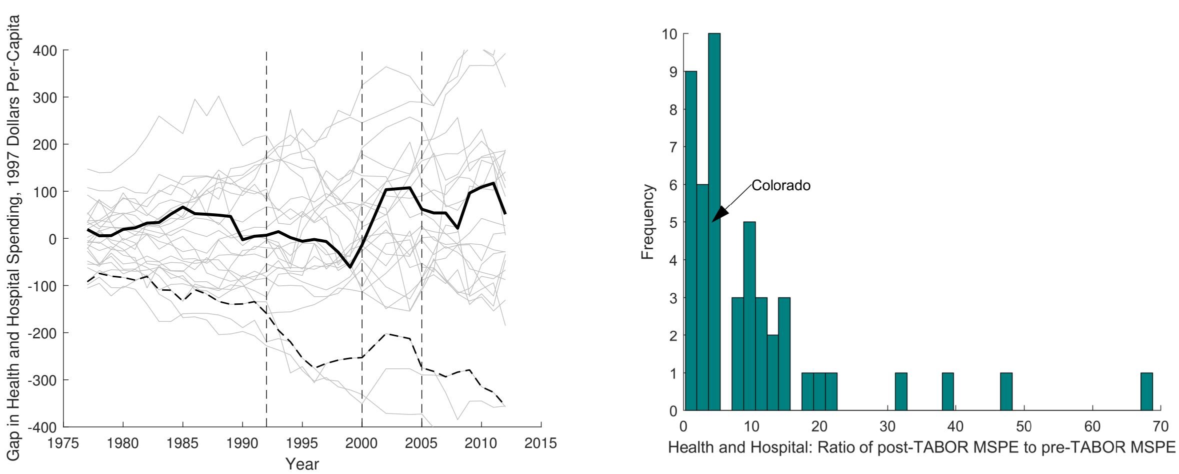
Note. Fixed weight approach with state weights obtained from simultaneous estimation of expenditures and taxes (displayed in Figures 1 and 2). Placebo states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE for the simultaneous tax and expenditure outcomes are excluded in the left-hand figures. All control states are included in the right-hand side. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Figure 5: Components of Spending

Panel A: K-12 Education Spending Per Capita



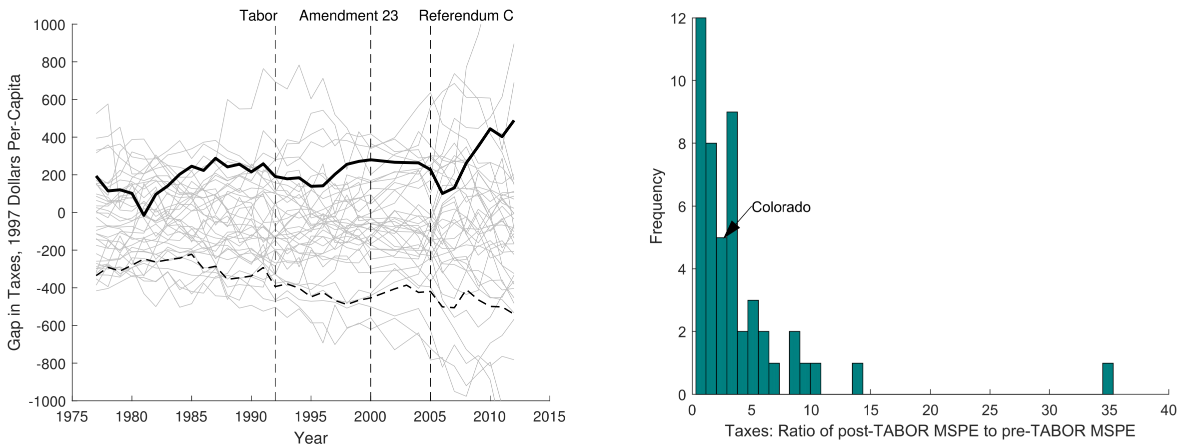
Panel B: Health and Hospitals Spending Per Capita



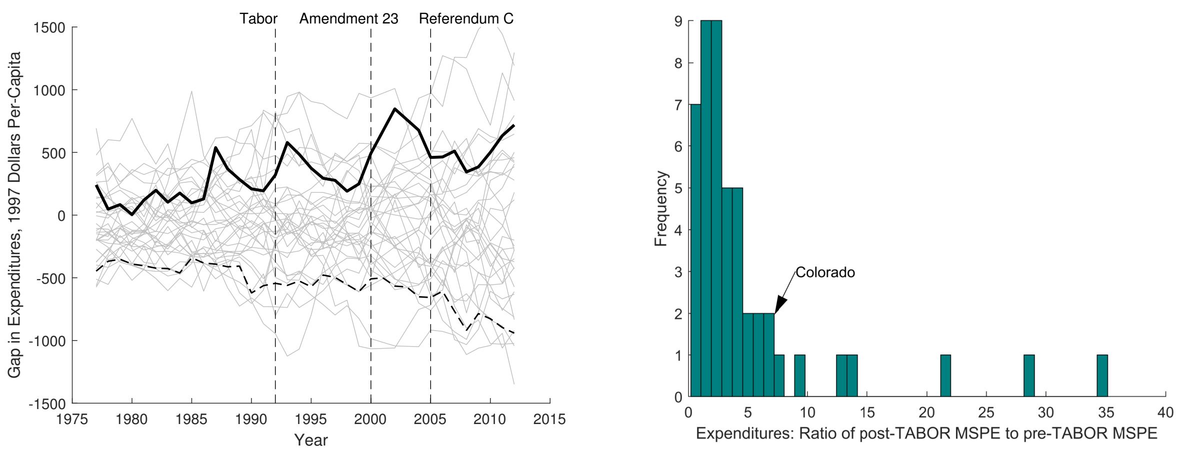
Note. Fixed weight approach with state weights obtained from simultaneous estimation of expenditures and taxes (displayed in Figures 1 and 2). Placebo states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE for the simultaneous tax and expenditure outcomes are excluded in the left-hand figures. All control states are included in the right-hand side. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Figure 6: Robustness Checks

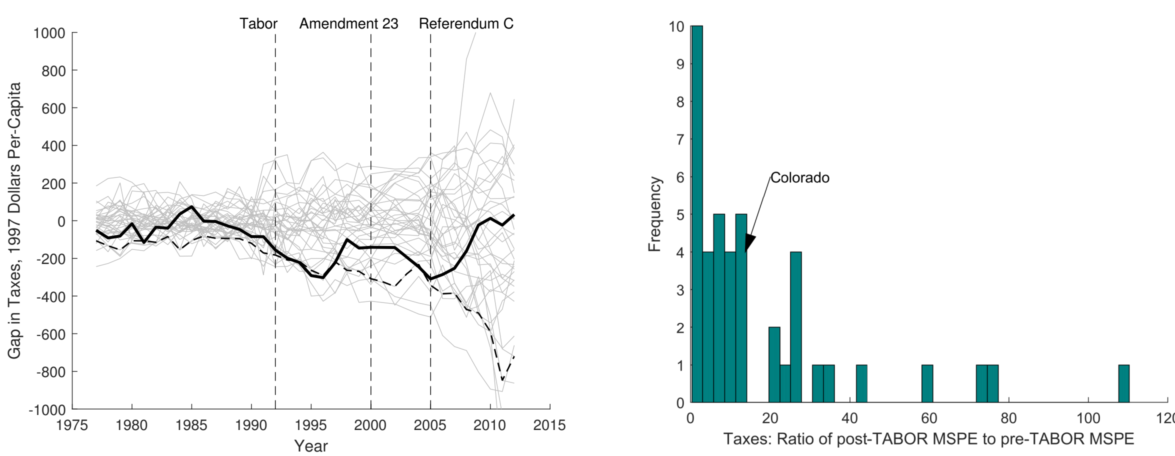
Panel A: Taxes, Computed Using Donor Pool Restricted to Equally Weighted Mountain Division States



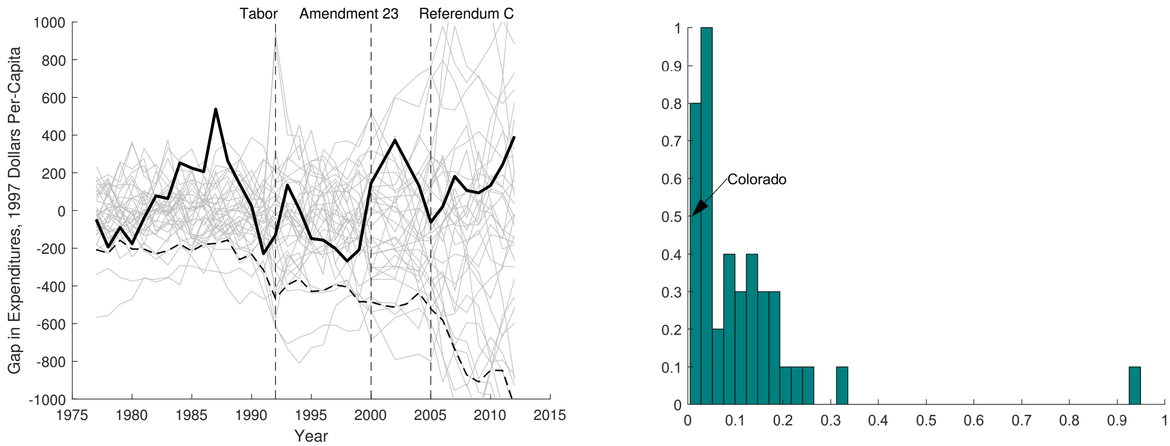
Panel B: Expenditures, Computed Using Donor Pool Restricted to Equally Weighted Mountain Division States



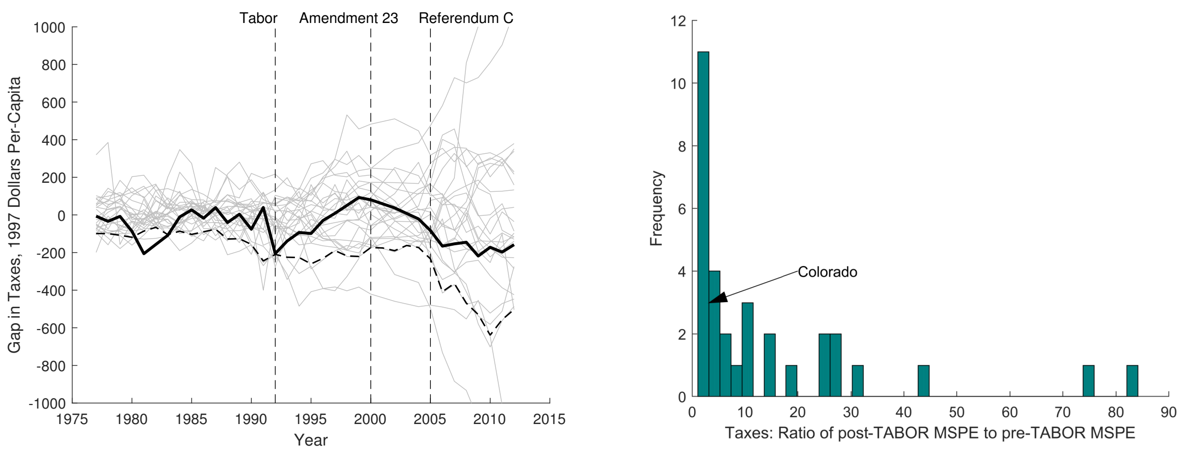
Panel C: Taxes, Computed Using Donor Pool that Excludes Bordering States



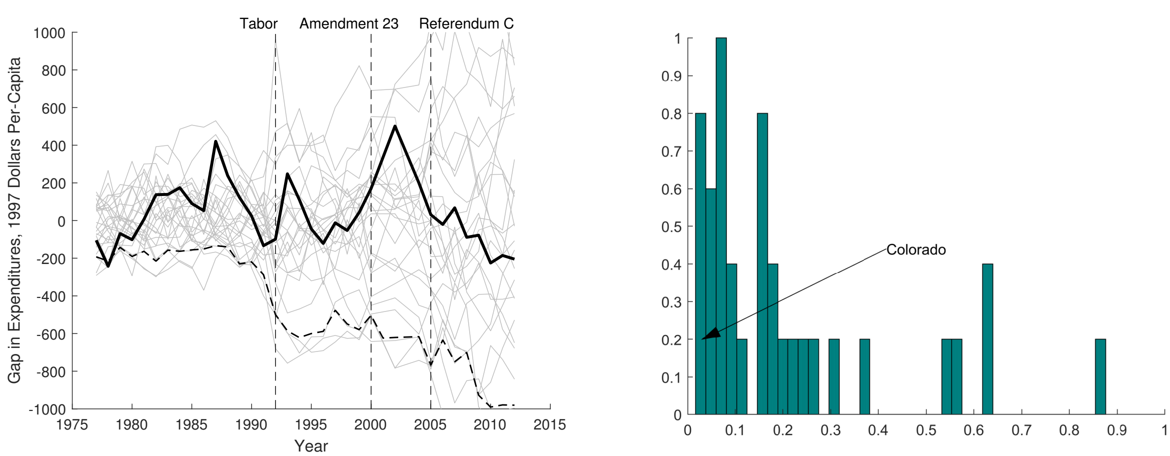
Panel D: Expenditures, Computed Using Donor Pool that Excludes Bordering States



Panel E: Taxes, Computed Using Donor Pool that Excludes States that Enacted TELs in 1987 or Later

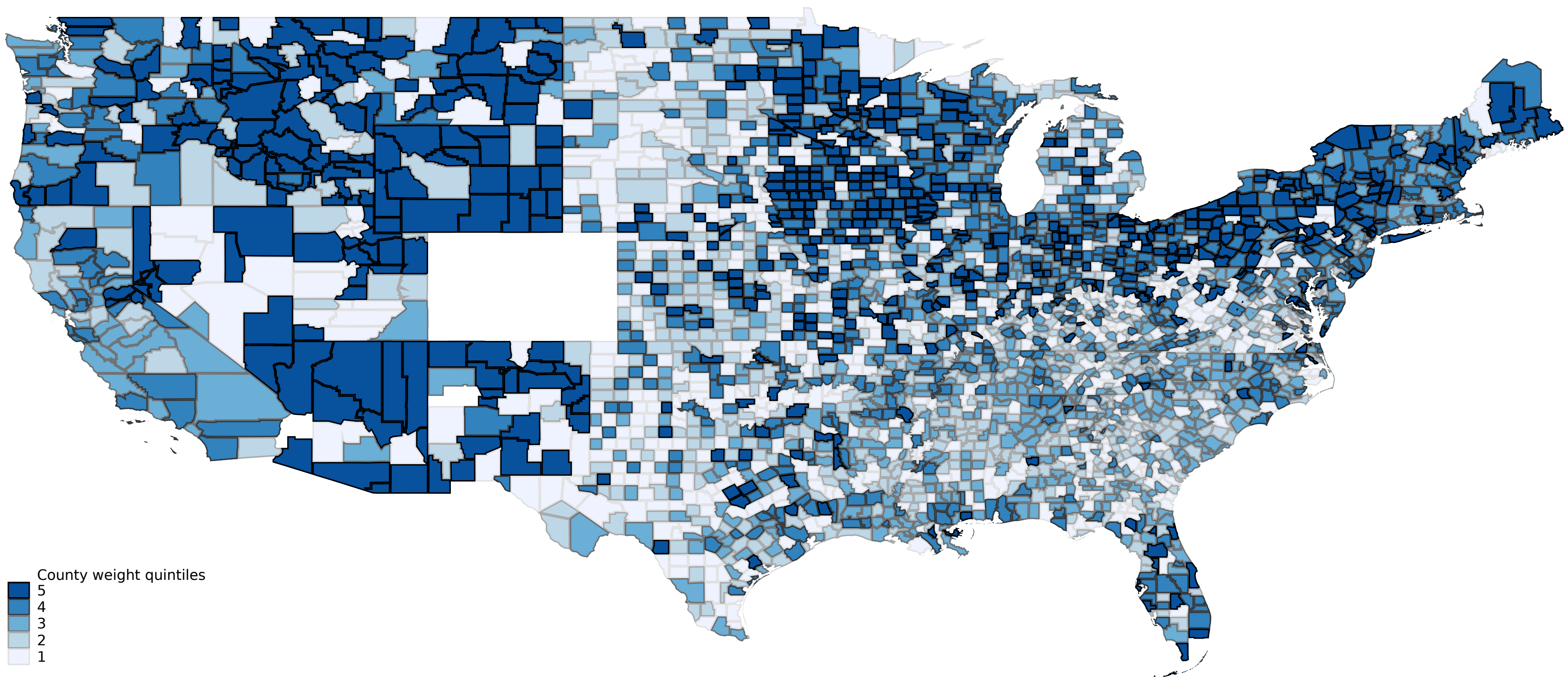


Panel F: Expenditures, Computed Using Donor Pool that Excludes States that Enacted TELs in 1987 or Later



Note. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in the left-hand figures. All control states are included in the right-hand side. In Panels A and B the donor pool is restricted to mountain division states and each of these states has an equal weight in W. In Panels C and D the states which physically border Colorado are excluded from the donor pool. In Panels E and F any state enacting a TEL in 1987 or later is excluded from the donor pool. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

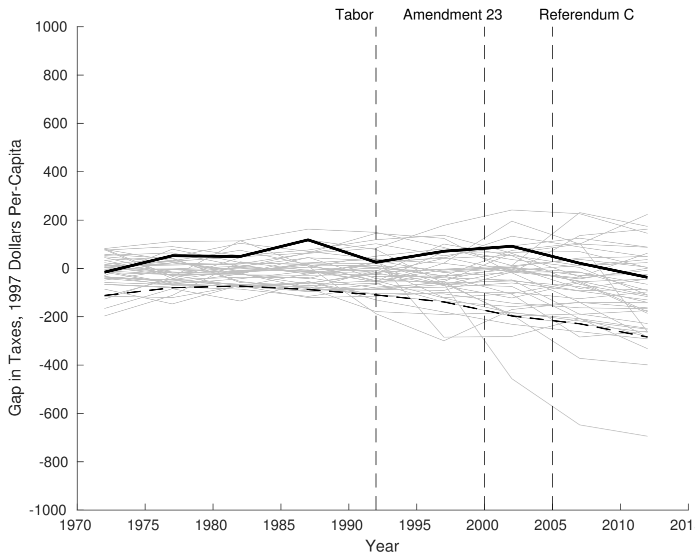
Figure 7: Weights on Non-Colorado Counties in the Synthetic Control County-Based Estimator



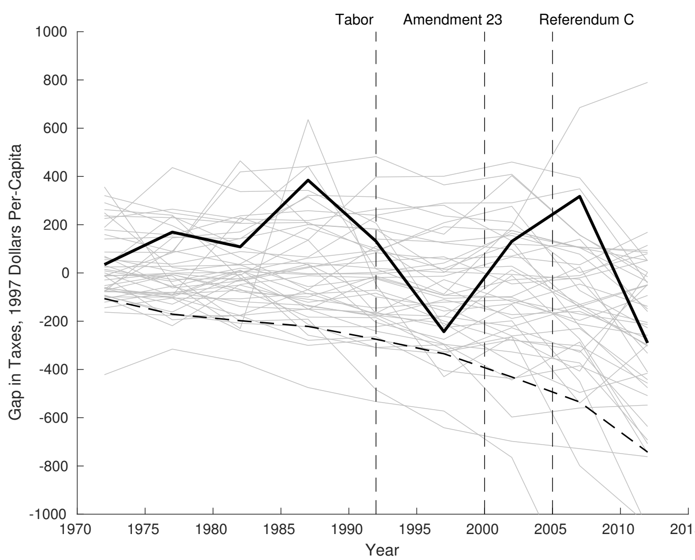
Note: Each non-Colorado county receives a non-negative weight in the synthetic control county-level estimator. These weights reflect both the simultaneous method synthetic control weights w_c^*, where c indexes counties (see section 2.1), and the population weighted averaging used to generate the single state-level treatment effect. Alaska and Hawaii counties–not displayed–are assigned infinitesimal weight by the procedure.

Figure 8: Synthetic Control County-Based Estimator

Panel A: Taxes Per Capita



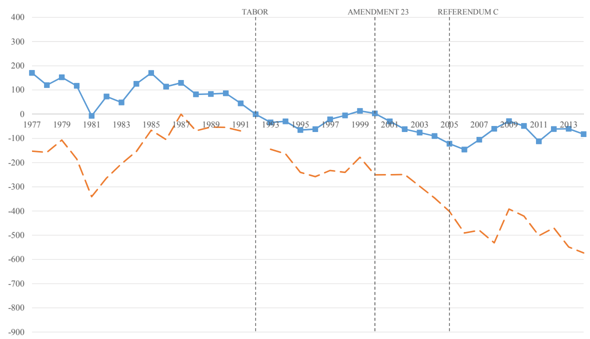
Panel B: Expenditures Per Capita



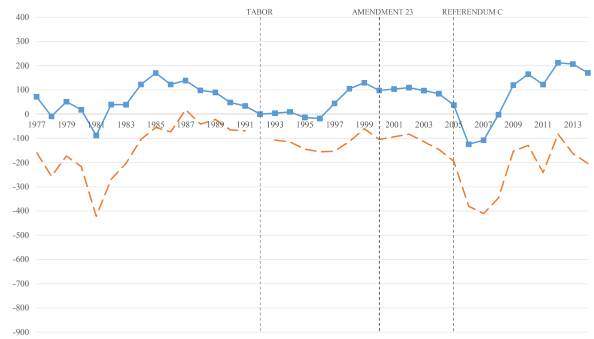
Note. Estimates calculated county-by-county using the simultaneous weight synthetic control method and then averaged to the state level using population weights. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded from the set of inference bands. The Census data on taxes and expenditures aggregated to the county-level are only available in years in which a census of all governments is taken; this occurs in years ending in 2 and 7. Correspondingly, the tax and expenditure results displayed here only take on values in years ending in 2 and 7. In both panels, the dashed line displays the lower bound 90% percent confidence limit (one-sided test).

Figure 9: Taxes and Expenditures Per Capita, Regression-Based Dynamic DD Approach

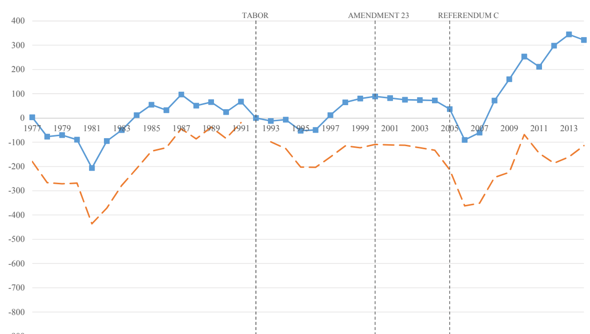
Panel A: Taxes, Year F.E.



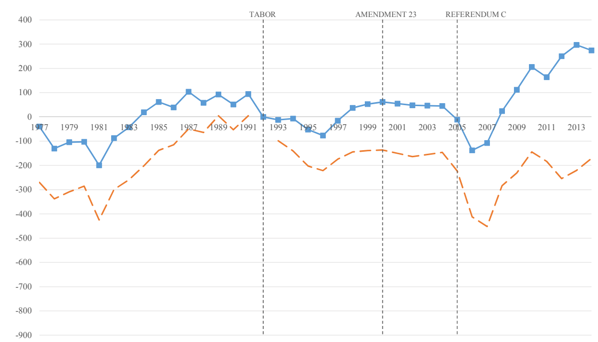
Panel B: Taxes, Year-Region F.E.



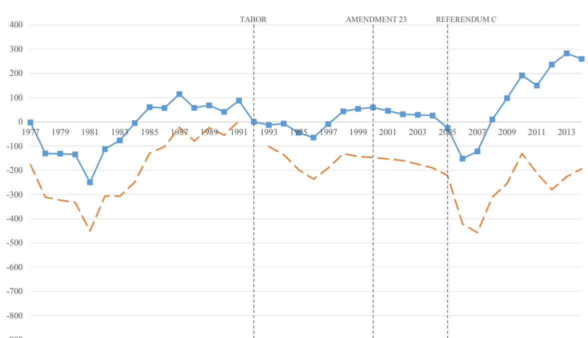
Panel C: Taxes, Year-Division F.E.



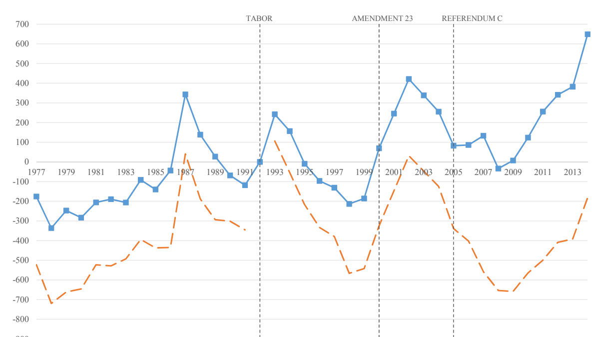
Panel D: Taxes, Year-Div. F.E., Controls 1



Panel E: Taxes, Year-Div. F.E., Controls 2



Panel F: Expenditures, Year-Div. F.E., Controls 2



Note. The lines with squares display the β vector coefficient estimates from equation (8). The additional line displays the 90% Conley and Taber (2011) lower confidence limit (one-sided test). The dependent variable is real tax collections per capita in panels A - E and real expenditures per capita in panel F. Panel D includes indicator variable controls for the presence of non-TABOR fiscal rules: general revenue limits, statutory expenditure limits, constitutional expenditure limits, and tax legislative supermajority requirements (labeled as "Controls 1"). Panels E and F allow the effect of these fiscal rules to vary with time relative to their implementation (labeled as "Controls 2") – see the text of section 4.4.3 for additional information. The sample includes all states other than AK and WY (and also excludes DC) and runs from 1977 to 2014.

Table 1: Summary Statistics

	Colorado	Synthetic Colorado	All Donor Pool States
A. Predictors of Budget Outcomes			
Census Division			
Mountain Division	*1.00*	*0.62*	*0.13*
Political Variables			
Republican Lower House	*1.00*	*0.76*	*0.22*
Republican Upper House	*1.00*	*0.90*	*0.26*
Democratic Governor	*1.00*	*0.62*	*0.59*
Sectoral Mix			
% of GSP from Manufacturing	*13%*	*15%*	*20%*
% of GSP from Finance	*18%*	*16%*	*15%*
% of GSP from Mining	*4%*	*3%*	*3%*
State Unemployment Rate	6.1	5.9	6.6
Income per capita	24176	22026	22044
Age Profile:			
Pop. under 18	*29.3%*	*31.1%*	*29.7%*
Pop. over 64	*8.7%*	*10.5%*	*11.2%*
Education Profile:			
Pop. graduated high school	*76.4%*	*72.0%*	*65.0%*
Pop. with bachelor degree	*21.9%*	*17.0%*	*15.2%*
Population Growth (thousands)	49.1	45.5	47.8
B. Per-Capita Tax Collections			
Taxes per-capita 1980	1787	1801	1696
Taxes per-capita 1985	2101	2037	1957
Taxes per-capita 1990	2318	2323	2257
C. Per-Capita Expenditures			
Expenditures per-capita 1980	3340	3443	3245
Expenditures per-capita 1985	3982	4061	3641
Expenditures per-capita 1990	4531	4698	4317

Note. All cells display means. For the sectoral mix, census division, political variables and age profile categories only selected variables are displayed. All dollars are in constant 1997 dollars.

Table 2: State Weights in the Synthetic Colorado

State	W Weights
AZ	0.31
KS	0.23
NV	0.10
NH	0.05
NY	0.09
UT	0.22
Total	1.00

Note. State weights underlying the synthetic Colorado produced by simultaneous weight method with taxes and total government expenditures used as the outcome variables.

Table 3: State Government Revenue Changes Due to Policy Changes

Year	Colorado	Synthetic Colorado
1996	0	-52
1997	-34	-129
1998	-127	-180
1999	-153	-219
2000	-231	-222
2001	-250	-240

Note. Changes in real dollars per capita. For Colorado these changes represent TABOR rebates plus the cumulative net change in revenue due to changes in state government tax policy in fiscal years 1996 through 2001. For the synthetic state they represent the cumulative net change in revenue due to changes in state government tax policy. Changes in revenue due to changes in tax policy are obtained from the fall editions of the *Fiscal Survey of the States* produced by the National Association of State Budget Officers (NASBO).

Table 4: Relative Pre-TABOR Fit of Different Estimation Approaches

	Optimizing Over Taxes	Optimizing Over Expenditures	Simultaneous Optimization
Taxes MSPE	10,807 (Floating Weight)	206,072 (Priority Weight w/ Exp. Prioritized)	17,525 (Simultaneous Weight)
Expenditures MSPE	436,510 (Priority Weight w/ Taxes Prioritized)	141,287 (Floating Weight)	164,472 (Simultaneous Weight)

Note. The table displays the mean squared predicted error (MSPE) for pre-TABOR period for the three estimation approaches – floating weight (i.e. the standard procedure), priority weight and simultaneous weight.

6 Appendix

6.1 TABOR Rebates

TABOR does not specify how excess revenue should be rebated to voters; it only requires a "reasonable method" be used. The state legislature chose to rebate excess revenue through a set of contingent tax credits available through the state's income tax system. The availability of the credits was contingent on revenue collections being in excess of the TABOR limits in the previous tax year. The magnitude of the credits was a function of the amount of excess revenue.

In the first two years of the rebates—fiscal years (FY) 1997 and 1998—only a single contingent tax credit was available. The credit was labeled as a "sales tax refund", but was issued as a function of a taxpayer's federal adjusted gross income (AGI) and filing status (i.e. higher AGI households received larger credits). The credit amounts were set each year to return all of the excess revenue collected in the previous year.

In FY 1999, the state introduced two additional TABOR contingent tax credits – a state earned income tax credit and a refund for property tax payments made by businesses – for a total of 3 contingent credits. In FY 2000 there were a total of 9 contingent credits and in FY 2001 there were 16 contingent credits. These credits were progressively triggered at different levels of excess revenue. In principal, as the amount of excess revenue rose, additional credits were triggered. In practice, all credits available in a given year were triggered from FY 1997 - FY 2001. From FY 1999 - FY 2001 the sales tax rebate was the residual category in the sense that the amount of the credit was set each year to ensure that all of previous year's excess revenue was returned to taxpayers. The other credits were set at fixed amounts by the legislature. The sales tax credit was the most significant credit over this period, equaling around 88 percent of total credits issues in FY 1999, 72 percent of total credits in FY 2000 and 64 percent of total credits in FY 2001.

6.2 Sensitivity of Estimates to Varying the X Vector

In this Appendix section we assess the sensitivity of the simultaneous method synthetic control estimates—displayed on Figures 1 and 2—to variation in the vector of observable fiscal predictors X. To do so, we estimate the simultaneous weight procedure over every combination of the X predictor set that contains at least two variables. Census division, political variables, sectoral mix, and age profile are sampled as groups—see Table 1 which displays all major categories of the X set. Overall, 506 iterations are estimated.

Figure A-6 displays the interquartile range of these estimates, as well as the median estimate. Panel A displays the results for taxes per capita and panel B displays the results total expenditures per capita. In both cases, the interquartile range is fairly narrow and has a contour very similar to that of the estimates which use all of the X predictors on Figures 1 and 2. Thus, we conclude that the synthetic control estimates are not overly sensitive to permutations in the X vector.

6.3 Non-TABOR Stringent Fiscal Rules

In this appendix section we discuss how we define non-TABOR stringent fiscal rules. We base our fiscal rule classification on Resnick (2002) and the comprehensive listing of TELs maintained by the National Conference of State Legislatures (Waisanen, 2010). First, we categorize TELS as restricting either revenue or expenditures. We then exclude revenue limits which cap only a specific form of revenue (e.g. property taxes) because they can be easily evaded by increasing other forms of revenue (Kousser et al., 2008a) and thus are not strict. We exclude expenditure limits which cap spending based on projected revenues on the grounds that they are unlikely to bind and are thus not strict. Such limits can be easily evaded by manipulating the prospective revenue projections (Resnick, 2002). Next, we categorize the TELs as being constitutional limitations or statutory limitations – constitutional limitations are typically viewed as being substantially more restrictive because it is more difficult, or impossible, for the state legislature to cancel or adjust the limitations as compared to a rule put into place by statue (Resnick, 2002; Kousser et al., 2008a; New, 2010b). Thus, we categorize TELs into four groups. That said, there is only a single state which enacted a statutory general revenue limit within our sample period (all other general revenue limits were constitutional). As as result, in practice we use three categories of strict TELs. We also collect data on legislative super-majority requirements for tax increases. For these four types of fiscal rules–general revenue limit, statutory expenditure limit, constitutional expenditure limit, and tax supermajority requirement – we collect the date of enactment for any rule put into place in 1977 (the start of our sample period) or later.

6.4 Alternative Outcome Variables

We have measured budget outcomes exclusively in per capita terms. An alternative is to use the log of the budget outcomes or to normalize the outcomes by personal income. We explore these alternatives in this Appendix section.

Figure A-7 displays the results of using the simultaneous weight method with log taxes and log expenditures as the outcomes. The results are quite similar to the per capita tax and expenditure results on Figures 1 and 2.

Turning to normalizing the outcome variables by personal income, we strongly prefer the per capita approach because its interpretation involves fewer assumptions. For instance, a fall in per capita spending provides a relatively clear indication that the provision level of public goods has fallen. In contrast, concluding that a fall in spending as a share of income represents a reduction in public goods requires the implicit assumption that the income elasticity of government expenditures is greater than or equal to one or that it ought to be greater than or equal to one. We see little benefit in making such assumptions about voter preferences. Nonetheless, some may view the income normalization as informative because Colorado experienced robust economic growth in the 1990s (although as discussed in McGuire and Rueben (2006) this growth was in-line with that experienced by its Mountain state neighbors).

Figure A-8 presents results for taxes and total expenditures denominated by personal income. In panel A, the fixed weight approach is used with weights from our preferred results – the simultaneous weight estimates for *per capita* taxes and expenditures (displayed in Figures 1 and 2). The aim is to examine the outcome variables with a different denominator while holding the composition of the synthetic Colorado fixed to that used for our preferred results. The figures reveal that Colorado has unusual preferences as the treatment effect is near the lower envelope of placebo states in both the pre and post-TABOR periods: Relative to its synthetic state, with which it shares very similar per capita tax collections, Colorado has unusually low taxes and expenditures as a share of income. However, expenditures display a relatively flat trend over the entire sample period, providing no evidence of a TABOR effect.

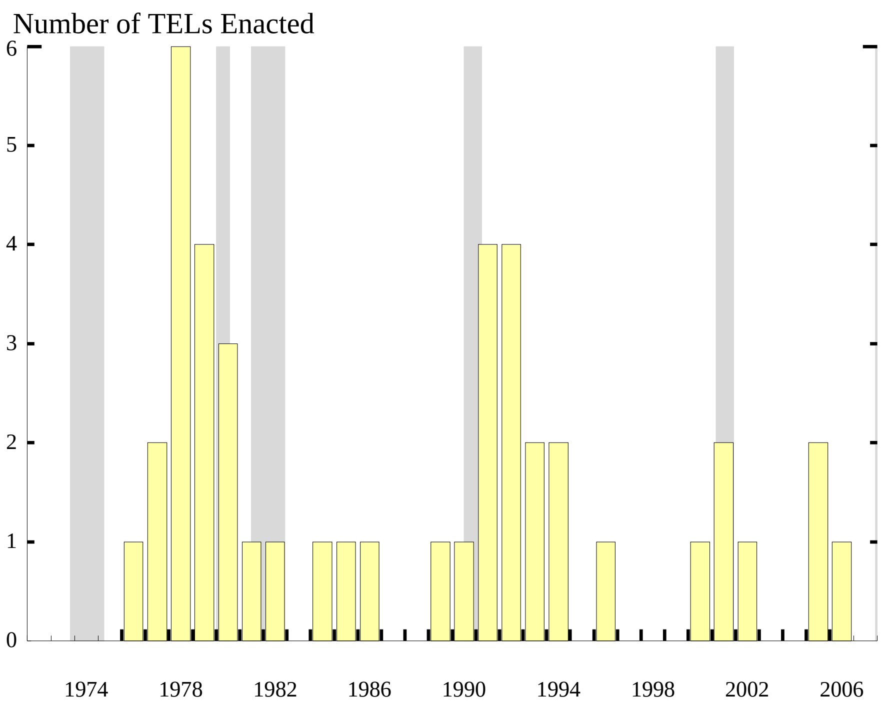
Taxes display a flat trend over the post-TABOR period. There is a downward tilt to taxes in the pre-TABOR period, but as this trend precedes the policy it is unlikely to have been caused by TABOR.

Panels B and C use the simultaneous weight method with taxes and expenditures as a share of personal income as the outcome variables. With regards to taxes, there are two important aspects of the results. First, the synthetic cohort procedure is much less successful at constructing a counterfactual for Colorado when taxes are denominated by personal income as opposed to population: The pre-treatment fit is relatively erratic in the left-hand side of Panel B. Second, although there is a downward tilt to the tax treatment effect, the downward slope precedes TABOR. This can be seen quite clearly in Panel B of Appendix Figure A-5 which replicates these tax results with the addition of a trend line. The trend line reveals that the treatment effect is nearly perfectly linear from 1987 to 1997, a period roughly divided in half by the 1991 passage of TABOR. In order to conclude that TABOR *caused* the downward tilt, the change in slope would need to occur at the time of TABOR's enactment or after, not before.

The expenditure treatment effect also displays a downward tilt from 1987 forward, but bounces back up around the time of TABOR enactment before declining again through the mid-1990s. In several years in the mid-1990s it is outside the lower bound of the placebo state treatment effects (although the alternative inference in the right of panel C suggests that the null of no TABOR effect cannot be rejected). Panel C of Appendix Figure A-5 presents analogous results for current expenditures—total expenditures minus capital expenditures, interest on debt, assistance payments and insurance benefits—which have lower annual volatility.[44] With the removal of capital expenditures, which are often quite lumpy from year-to-year, the procedure has a bit more success at forming a synthetic Colorado that matches the actual Colorado in the pre-Tabor period with respect to expenditures. Notably, there is a clear downward trend in the operating expenditure treatment effect which predates TABOR, casting significant doubt on the hypothesis that TABOR caused a decline in expenditures.

[44]These are estimated with the simultaneous weight approach with taxes and operating expenditures denominated by personal income as the outcome variables. The tax results are extremely similar to those in panel B of Figure A-8 and are available from the authors upon request.

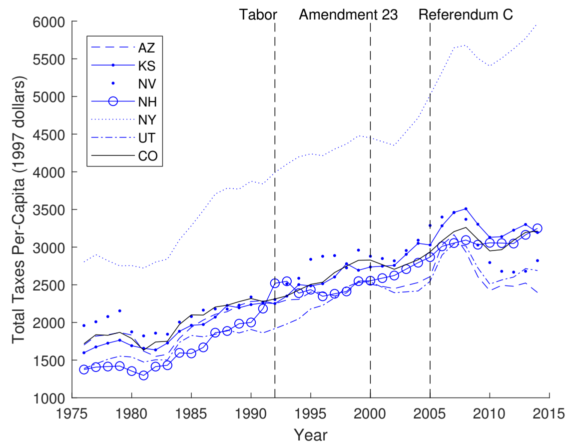
Figure A-1: Timing of TEL Enactment



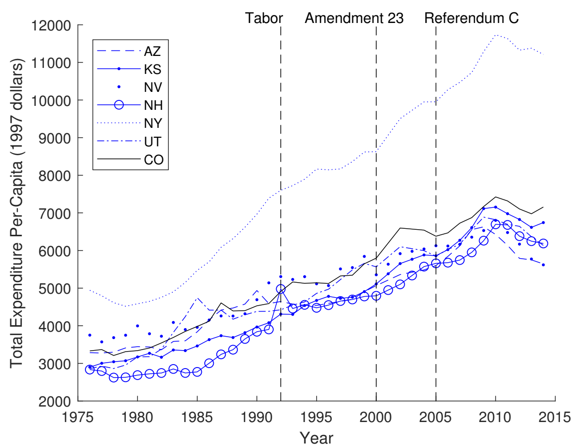
Note. Some states have passed more than one TEL over the period displayed. Source. Poterba and Rueben (1999) and Waisanen (2010).

Figure A-2: Taxes and Expenditures in Colorado and the Synthetic Colorado States

Panel A: Taxes Per Capita



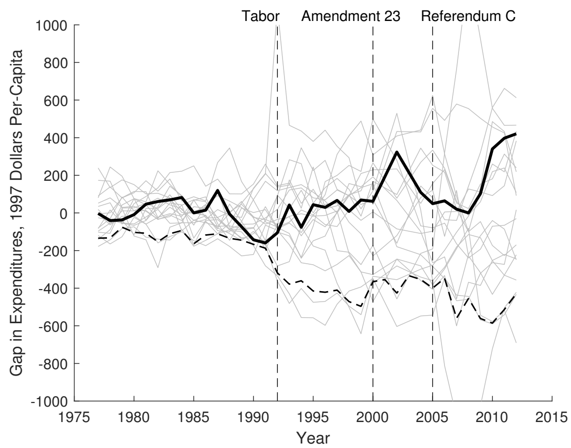
Panel B: Total Expenditures Per Capita



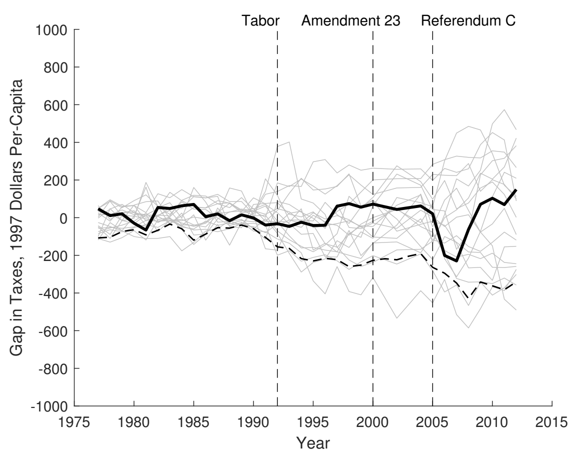
Note. The graphs display the state-level data upon which Panels A of Figures 1 and 2 are based.

Figure A-3: Current Expenditures Per Capita and Taxes Per Capita

Panel A: Current Expenditures Per Capita



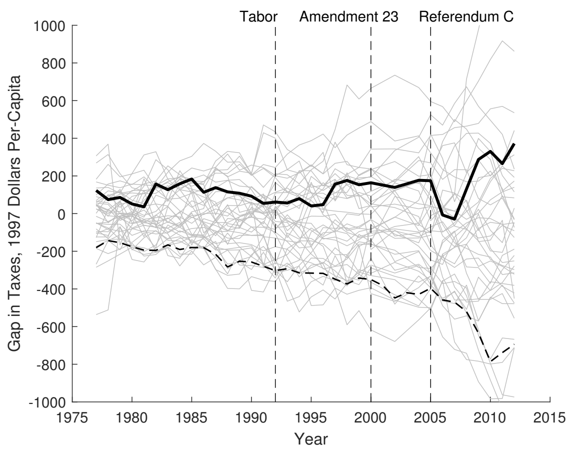
Panel B: Taxes Per Capita

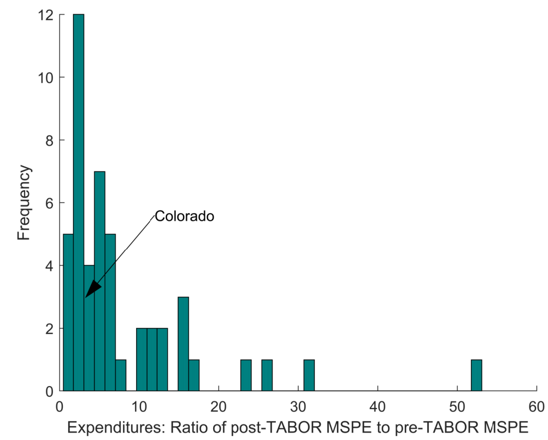


Note. Simultaneous weight method. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in the figures in Panels A and B. The dashed line displays the lower bound 90% percent confidence limit (one-sided test).

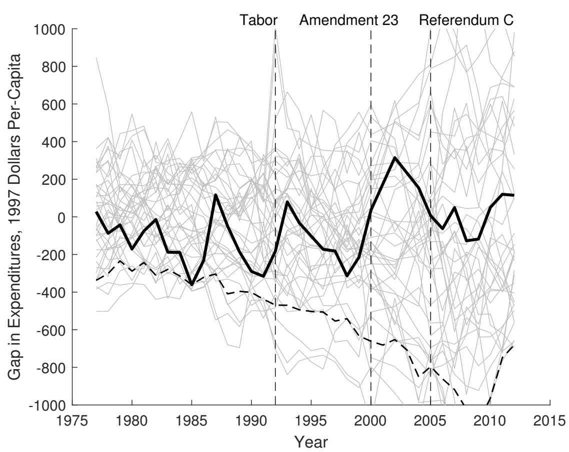
Figure A-4: Priority Weight Method

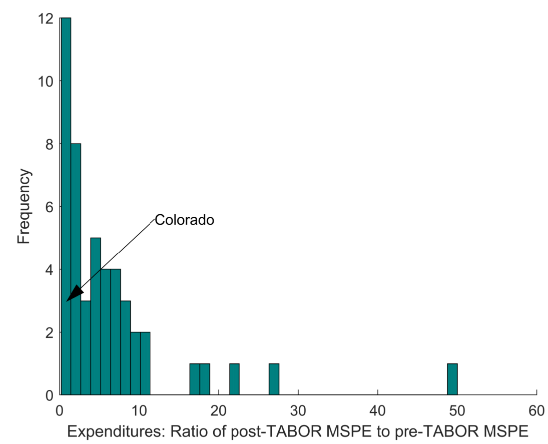
Panel A: Taxes Per Capita





Panel B: Total Government Expenditures Per Capita

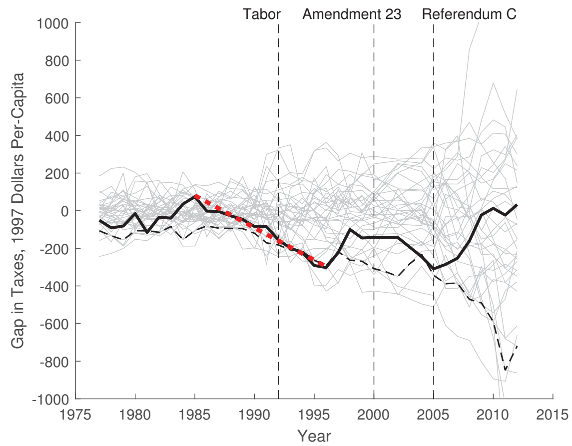




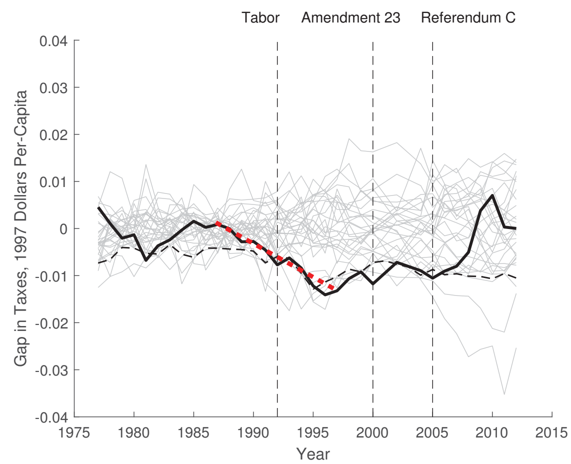
Note. Priority weight method. In Panel A the state weights are obtained by executing the standard synthetic control methodology with total government expenditures per capita as the outcome. These expenditure results are displayed in Panel B of Figure 3. In Panel B the state weights are obtained by executing the standard synthetic control methodology with taxes per capita as the outcome. These tax results are displayed in Panel A of Figure 3. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE *for the outcome variable used to generate the weights* are excluded in the left-hand figures. (I.e. the states displayed as inference bands in the expenditure results in Panel B of Figure 3 are the same states used to produce the inference bands on Panel A of this figure.) All control states are included in the right-hand side figures. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Figure A-5: Trend Lines

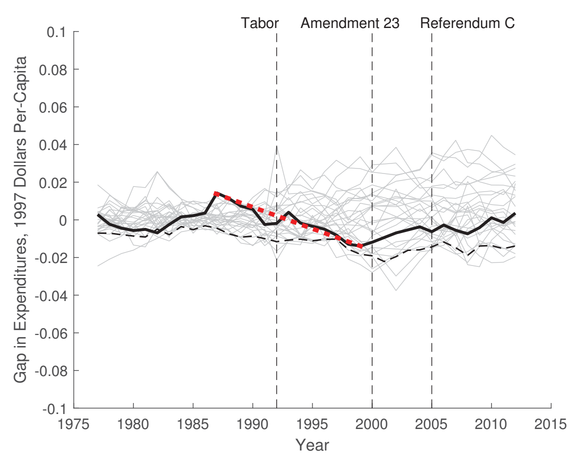
Panel A: Taxes with Bordering States Excluded from Donor Pool with Trend Line



Panel B: Taxes as Share of Personal Income with Trend Line



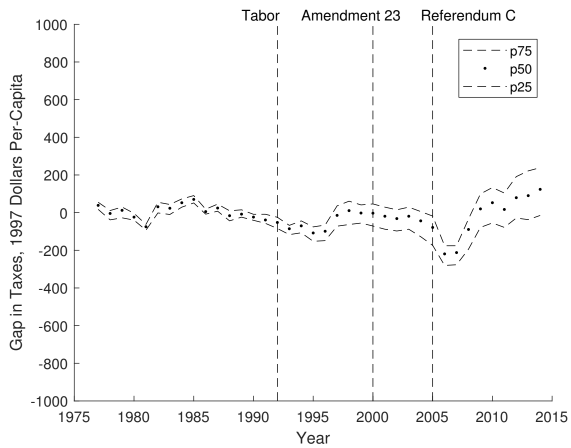
Panel C: Current Expenditures as Share of Personal Income with Trend Line



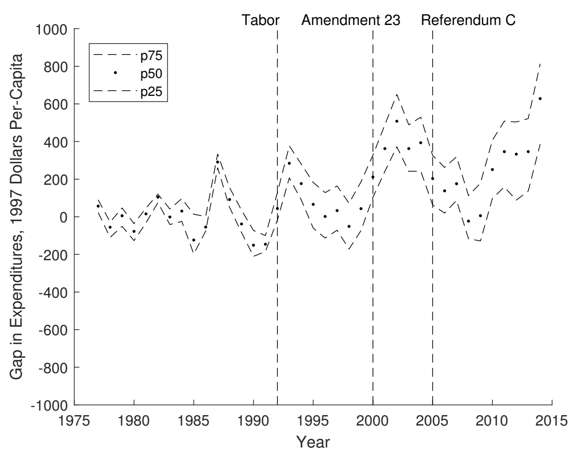
Note. Panel A: simultaneous weight method with taxes per capita and expenditures per capita as the outcome variables and border states excluded from the donor pool (replicates panel C of Figure 6). Panel B: simultaneous weight method with taxes as a share of personal income and total government expenditures as a share of personal income as the outcome variables (replicates panel B of Figure A-8). Panel C: simultaneous weight method with taxes as a share of personal income and operating expenditures as a share of personal income as the outcome variables. In all panels, the dashed line displays the lower bound 90% percent confidence limit (one-sided test).

Figure A-6: Sensitivity of Estimates to Varying the X Vector of Predictor Variables

Panel A: Taxes Per Capita



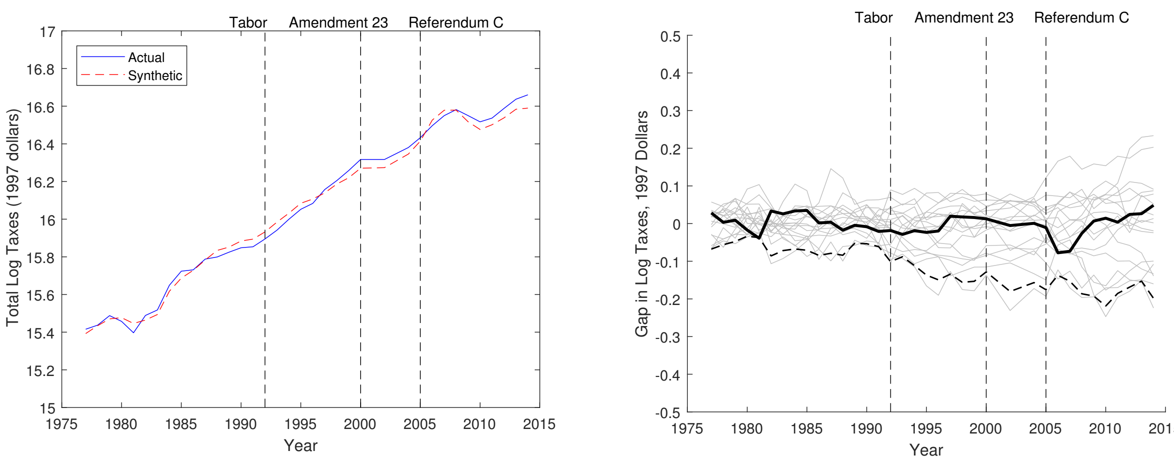
Panel B: Total Expenditures Per Capita



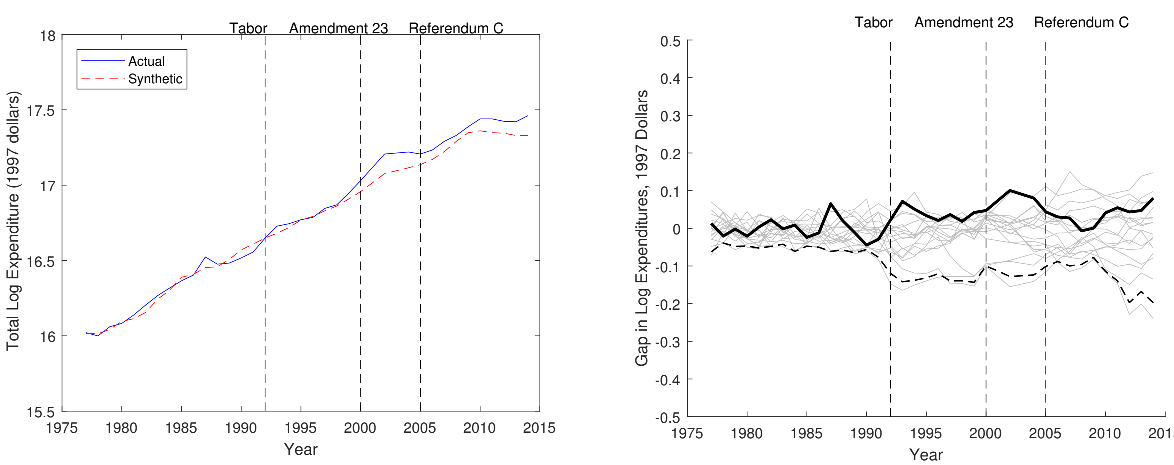
Note. The simultaneous weight method is estimated separately for all possible combinations of the X fiscal predictor set which contain two or more variables. Census division, political variables, sectoral mix, and age profile are sampled as groups–see Table 1. In total, 506 iterations are estimated. The figures plot the interquartile range of these estimates, as well as the median estimate.

Figure A-7: Simultaneous Weight Approach with Outcomes in Logs

Panel A: Log Taxes



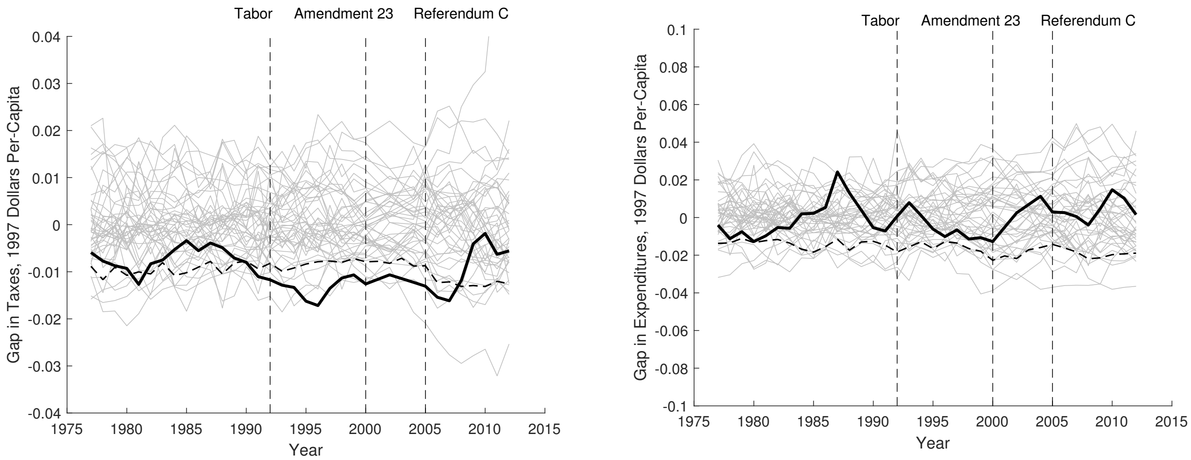
Panel B: Log Total Government Expenditures



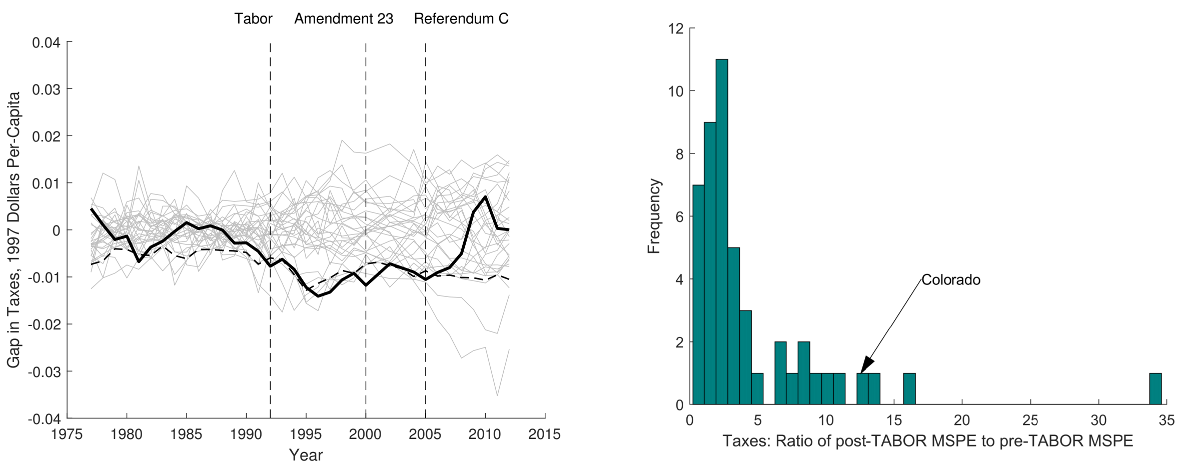
Note. Control states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in the right-hand figures. The dashed line in the right-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Figure A-8: Taxes and Expenditures as a Share of Personal Income

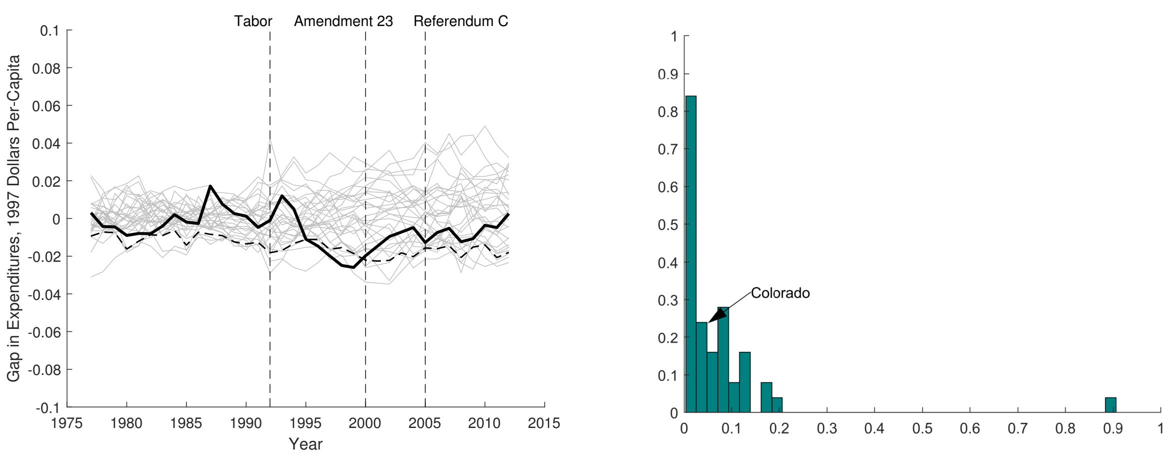
Panel A: Taxes and Expenditures as Share of Personal Income, Fixed Weight Method



Panel B: Taxes as Share of Personal Income, Simultaneous Weight Method



Panel C: Expenditures as Share of Personal Income, Simultaneous Weight Method



Note. Panel A uses the fixed weight method with state weights obtained from simultaneous estimation of expenditures and taxes *per capita* (displayed in Figures 1 and 2). Placebo states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE for the simultaneous tax and expenditure per capita outcomes are excluded. Panels B and C use the simultaneous weight method with taxes as a share of personal income (panel B) and total government expenditures as a share of personal income (panel C) as the outcome variables. Placebo states with pre-TABOR MSPE greater than 5 times Colorado's pre-TABOR MSPE are excluded in the left-hand figures of Panels B and C. The dashed line in the left-hand figures displays the lower bound 90% percent confidence limit (one-sided test).

Table A1: TABOR Literature Review

Author	Publication or Organization	Conclusions in Regards to TABOR's Effect on Taxes and Spending
		A. Journal and Book Publications
Billings, Stephen and Deborah Carroll (2012)	Growth and Change	Counties and municipalities where debrucing initiatives failed were no more likely to create special districts (autonomous entities that have the power to tax, spend, and issue debt) than counties and municipalities where debrucing initiatives were successful. In other words, being bound by TABOR had no effect on the creation of special district governments.
Brown , Tom (2000)	Public Budgeting and Finance	TABOR constrained municipal budgets. The effects were not uniform across
James, Franklin and Allan Wallis (2004)	Public Budgeting and Finance	TABOR has created significant reductions in the growth rate of spending; the revenue limits have also led to temporary and permanent tax cuts. Numerous municipalities have received voter approval for "de-Brucing," a process which allows them to retain excess tax revenues.
Kousser, Thad; McCubbins, Mathew; and Kaj Rozga (2006)	Fiscal Challenges: An Interdisciplinary Approach to Budget Policy, eds. Garrett,	TABOR reduced the size of Colorado's government compared to states without TELs; however, in 2005, voters passed a five-year suspension of the spending restrictions.
Kousser, Thad; McCubbins, Mathew; and Ellen Moule	State Politics & Policy Quarterly	Colorado was the only state with a TEL that reduced per-capita spending; nevertheless, TABOR has been undermined with the 2005 passage of the "timeout for TABOR".
Martell, Christine and Paul Teske (2007)	Public Administration Review	TABOR reduced and the size of Colorado’s state government relative to the economy and other western states with a TEL.
New, Michael (2010)	State Politics & Policy Quarterly	TABOR successfully limited the growth of state government in the 1990s, but the recession in 2001 created enough pressure on the state budget to lead policymakers to temporarily suspend some of TABOR's provisions.

Table A1: TABOR Literature Review (cont.)

Author	Publication or Organization	Conclusions in Regards to TABOR's Effect on Taxes and Spending
	B. Selected Policy Pieces and Unpublished Papers	
Atkins, Chris (2005)	Tax Foundation	TABOR's limit on revenue growth led to more stable streams of revenue and reduced the size of Colorado's budget deficits. TABOR has not led to a reduction in quality for programs such as health care and education.
Bradley, David (2005)	Center on Budget and Policy Priorities	TABOR caused major state programs to be underfunded (e.g. higher education and children's health care).
Citizens Budget Commission (2010)	Citizens Budget Commission	TABOR's revenue limits created budget problems during the 2001 recession and the ensuing economic recovery, as the revenue limits were tied to the lower recession
Greenwood, Daphne and Tom Brown (undated)	Center for Colorado Policy Studies	As municipal governments increasingly depend on sales tax revenues instead of property taxes, their revenue streams have becomes less stable; this situation is exacerbated by TABOR's restrictions on spending growth.
Hedges, Carol (2003)	Bell Policy Center	Programs that are not affected by the legislative budgeting process, such as Medicaid and Corrections, were shielded from budget restrictions, forcing other programs, such as Education and Health, to bear a disproportionate share of the spending restrictions. TABOR has also limited the potential for state spending to return to pre-recession
Lav, Iris and Erica Williams (2010)	Center on Budget and Policy Priorities	The spending restrictions imposed by TABOR mean there are insufficient resources to fund state services, leading to declining quality of programs such as health and education. The growth in the corrections budget exacerbates the inadequacies in spending in other areas.
McGuire, Therese and Kim Reuben (2006)	Economic Policy Institute	TABOR did not have a statistically significant effect on personal income growth in Colorado. While there is some evidence that TABOR had a positive effect on employment growth in the 5 years immediately following its passage, these gains were not sustained in the long term. In other words, TABOR was not responsible for Colorado's economic growth in the 1990s.
New, Michael and Stephen Slivinski (2005)	Cato Institute	TABOR is not the cause of Colorado's budget shortfalls in the 2001-2002 recession; rather, Amendment 23, an education spending mandate; a severe drought; and the recession are to blame.

Table A1: TABOR Literature Review (cont.)

Author	Publication or Organization	Conclusions in Regards to TABOR's Effect on Taxes and Spending
		B. Selected Policy Pieces and Unpublished Papers
New, Michael (2014)	National Review	Between 1997 and 2002, Colorado had the most tax relief and economic growth in the U.S., with refunds of $3.2 billion during this five-year period. However, at the turn of the 21st century, Colorado ran into economic trouble, and in 2005, voters approved the suspension of TABOR's revenue limits for five years. The revenue limit returned in FY 2011, and tax revenues will likely exceed the TABOR-imposed cap in FY 2014 for the first time since the return of the revenue limit.
Policy Basics (2013)	Center on Budget and Policy Priorities	TABOR's "population-plus-inflation formula" does not provide sufficient growth in state spending to maintain services, because the segments of the population that require the most state services frequently grow faster than the aggregate state population and because the inflation measure used for TABOR (the CPI-U) grows more slowly than the cost of goods and services that the state provides. Further, under TABOR, several of Colorado's public services have fallen in the national rankings. Specifically, Colorado's position in the national rankings for elementary, secondary, and post-secondary education funding and the provision of full, on-time vaccinations for children has fallen, while the percentage of low-income children without health insurance doubled.