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

April 2015
Preliminary – Do Not Cite Without Permission

We are grateful to Carol Hedges, Director of the Colorado Fiscal Policy 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. 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 Shanna Rose for helpful comments.

The analysis and conclusions set forth are those of the authors and do not indicate concurrence by the Board of Governors of the Federal Reserve.
Fiscal rules attempt to alter budget outcomes by constraining policy makers. They have become one of the primary responses to fiscal policy failures such as large and persistent deficits and systemic overspending. In light of the many recent fiscal crises around the globe, they will almost certainly take on increasing importance in the years ahead. It is unclear, though, whether these rules cause a change in budget outcomes, induce government actors to evade the rules, or merely ratify the existing preferences of a jurisdiction’s voters and officials. We ask if fiscal rules are capable of constraining the size of government by examining tax and expenditure limitations—one of the most prevalent forms of fiscal rules in the U.S. and the form most directly aimed at restraining the size of government. We focus on the most stringent of these limitations: Colorado’s Taxpayer Bill of Rights (TABOR). TABOR applies to all sub-national levels of government, 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. Numerous other states have considered, or are considering, adopting similar policies. 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. In order to evaluate TABOR, we construct a synthetic Colorado to provide counterfactuals for the budget outcomes which would have occurred in the absence of the limit. We do so by extending the synthetic control methodology of Abadie and Gardeazabal (2003) and Abadie et al. (2010) to allow for the simultaneous examination of two outcome variables (taxes and expenditures). The methodology uses a rigorous, data-driven approach to construct the synthetic Colorado and inferential methods appropriate for studying an intervention in a single state are employed. Our results suggest that TABOR had no impact on the level of taxes or spending in Colorado and provide no support 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.
Representative democracies often produce poor fiscal outcomes such as large and persistent deficits. Such outcomes likely 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 tax 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 the U.S. states and tax and expenditure limitations both state and local governments in the U.S. In the wake of the recent spat of fiscal crises around the globe – the deficit crisis in Europe, the downgrading of U.S. government debt over deficit concerns, and the state and local government fiscal crisis in the U.S. – fiscal rules are likely to take on ever greater importance. Indeed, the EU significantly tightened its budget rules under the Fiscal Compact of 2012 and several U.S. states are considering new tax and expenditure limitations.

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 fail to bind and serve as nothing more than a signal of voter preference. For instance, tax and expenditure limitations 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 signal or ratify an existing preference over the size of government.

In this paper we ask if fiscal rules are capable of constraining the size of the public sector. 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. 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.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). 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

¹Most of the other fiscal rules in use the U.S., such as balanced budget requirements and debt limitations, primarily aim to achieve budgetary balance.
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 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.\(^3\) The press has presented a similar view. For instance, one of the country’s leading pundits has repeatedly extolled the virtues of TABOR (Will, 2005, 2011), as has the editorial board of one of the nation’s leading newspapers (Wall Street Journal, 2002, 2004).

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 extending the aforementioned synthetic control methodology to permit the simultaneous analysis of two outcomes (taxes and expenditures). Specifically, we construct a synthetic Colorado from a weighted combination of states other than Colorado’s economy.


\(^3\)Martell and Teske (2007) conclude that TABOR shrunk the size of state government expenditures relative to the size of the state’s economy. However, they also document that per-capita state expenditures in Colorado track those of Arizona both before and after TABOR (see their Figure 1) – a result broadly similar to the results in this paper.
orado (none of which has a TABOR-style TEL). 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 method has three primary advantages for our study. 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 (2008), the unobserved preference for small government manifests itself differently over the course of the business cycle.

Our results suggest that TABOR has no influence on 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. We therefore conclude that in the absence of TABOR, tax collections would have been nearly identical to those actually collected under TABOR. 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. Finally, it reflects buyers remorse, as voters relaxed, but did not repeal, the policy on several occasions.

These findings are 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 application and extension of the synthetic control method and Section 3 discusses the data. Section 4 presents our empirical findings which are then interpreted in section 5. Section 6 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, 2008). 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. Several papers have attempted addressed 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).

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.\textsuperscript{5} 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 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. 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 levied in the state. At the state government level, the cap applies not only to the General Fund, but also to most cash

\textsuperscript{5}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).
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. 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 (Baron et al., 2003). 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-

---

6See for example McGuire and Rueben (2006); Waisanen (2008); Resnick (2002).
level – population growth plus 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 the state income tax.

Significant 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 it 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 choose to forgo a portion of their TABOR refund in order to increase education spending. The

\(^7\)Discussions with multiple Colorado officials suggest that no systematic data is available on these rebates. In many cases, the limits were adhered to not by issuing rebates, but by lowering the millage rate of the property tax (Greenwood and Brown, 2000).
amendment also contained the implication that a higher share of the state budget would go to K-12 education over time.  

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

---

8 The discussion of Amendment 23 is a simplification which captures the key elements. See Legislative Council Staff (2001) for additional information.
period, it remained in effect and may have influenced budget decisions in the state.

2 Empirical Approach

The synthetic control method of Abadie and Gardeazabal (2003) and Abadie et al. (2010) 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). Such 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 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.9

We extend the synthetic control approach to a setting where there are two relevant outcomes (taxes and expenditures), both of which may be affected by the treatment. If the econometrician intends 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 and differences in the composition of the control group. We address this issue by altering the standard synthetic control approach to

---

9The 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 estimating the value of political connections (Acemoglu et al., 2010), the effect of Kyoto Protocol emissions targets (Almer and Winkler, 2011), the influence of natural disasters on economic growth (Cavallo et al., 2011) and the effects of California’s affirmative action ban on college racial integration (Hinrichs, 2012).
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

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

(1)

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

$$\hat{\alpha}_{J+1,t} = Y_{J+1,t} - \sum_{i=1}^{J} w^*_i Y_{i,t},$$

(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

$$W^* = \arg\min_{W} (X_1 - X_0W)'V(X_1 - X_0W)$$

(3)

where $X_1$ is a $(k \times 1)$ vector of pre-TABOR variables from Colorado which have predictive power for tax collections and $X_0$ is a $(k \times J)$ matrix of the same pre-TABOR predictors from the donor pool states.\(^{10}\) $V$ 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.

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

$$V^* = \arg\min_{V} (Z_1 - Z_0W^*(V))'V(Z_1 - Z_0W^*(V))$$

(4)

where $Z_1$ is a $(T_o \times 1)$ 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 $(T_o \times J)$ 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.

\(^{10}\)For example, we include census division indicators in our predictor vector because geographic differences in preferences for public goods have significant explanatory power for per-capita tax collections.
Stated differently, the method aims to produce a pre-TABOR match which reflects fundamental drivers of fiscal choice. 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

\[
Y_{i,t} = Y_{i,t}^0 + \alpha_{i,t}D_{i,t} \tag{5}
\]

where \(D_{i,t}\) is an indicator for treatment. Abadie et al. (2010) demonstrates that \(Y_{i,t}^0\) can be characterized as a factor model

\[
Y_{i,t}^0 = \delta_t + \Theta_tZ_i + \lambda_t\mu_i + \varepsilon_{i,t} \tag{6}
\]

Notice that the generality of this form allows the unobservable confounders, \(\mu_i\), to have a time-varying influence (i.e. \(\lambda\) is allowed to vary with \(t\)). Alternative methods for estimating equation (1), such as difference-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, 2008). 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. Differences in treatment effects across outcomes thus 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. Similarly, conclusions formed from jointly examining the two treatment effects may be problematic. 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 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 “fixed weight” method.

In order to overcome the drawbacks inherent in the floating and fixed 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:

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

(7)

where $E^1_{i,t}$ represents state and local expenditures under TABOR and $E^0_{i,t}$ represents expenditures in the absence of TABOR. 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

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

(8)

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

$$
V^* = \arg\min_V (\tilde{Z}_1 - \tilde{Z}_0 \tilde{W}^*(V))' (\tilde{Z}_1 - \tilde{Z}_0 \tilde{W}^*(V))
$$

(9)

where $\tilde{Z}_1$ is a $(2T_o \times 1)$ vector containing a $(T_o \times 1)$ vector of total taxes per capita in Colorado in each of the $T_o$ pre-TABOR years stacked on top of a corresponding $(T_o \times 1)$ vector of expenditures per capita in Colorado. The tax and expenditure variables are standardized such that they both have a variance and standard deviation equal to 1.\textsuperscript{11} Absent this

\textsuperscript{11}The outcome variables are standardized by dividing each outcome by its standard deviation in the given year.
standardization, the procedure would potentially place more implicit weight on one outcome than the other.

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 tests”. 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 relevant 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. (2011)

\[
p - value_t = \frac{\sum_{i=1}^{J} 1(\hat{\alpha}_{i,t}^{pl} < \hat{\alpha}_{J+1,t})}{\text{# of Placebo States}} = \frac{\sum_{i=1}^{J} 1(\hat{\alpha}_{i,t}^{pl} < \hat{\alpha}_{J+1,t})}{J} (10)
\]

where \( \hat{\alpha}_{J}^{pl} \) is the estimated treatment effect on tax collections of placebo state \( j \). A p-value for the effect on expenditures is analogous. The p-value ranks the Colorado treatment effect within the distribution of the placebo treatment effects.

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. The counterfactual explicitly includes the possibility of having an ordinary state-level TEL. This is appropriate 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.\footnote{Colorado’s ordinary TEL was enacted in 1977 and then strengthened in 1991.} Second, TABOR is widely considered a “super-TEL” that is substantially more effective than other TELs. Thus, it is appropriate to tests the efficacy of this “super-TEL” relative to ordinary TELs. Moreover, the 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.\footnote{The double difference is \([\text{post-Tabor Colorado taxes} - \text{post-Tabor synthetic Colorado taxes}] - [\text{pre-Tabor Colorado taxes} - \text{pre-Tabor synthetic Colorado taxes}]\). Any effect of the ordinary TEL on tax collections in the synthetic Colorado is differenced out. A bit more formally, the \(\mu_i\) term in equation (6) will control for the effect of an ordinary TEL in donor pool state \(i\).} We acknowledge, though, that a control state which enacts a TEL within the sample period could potentially introduce bias. We demonstrate that our results are robust to accounting for this possibility. Finally, we omit Alaska and Wyoming from the donor pool because they display extreme tax revenue volatility owing to the taxation of petroleum extraction. 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.

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
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.\footnote{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 of this situation, the income tax data reported to the Census 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.}

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 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.\footnote{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.}

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.\footnote{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.}

\footnote{These intercensal estimates are revised every ten years based on the Decennial Census. We use the unrevised estimates because these reflect the information available to policy makers at the time budget decisions were made. For instance, each year’s TABOR revenue cap is set using the unrevised estimates. We wish to capture the behavior of policy makers. Using the revised estimates would introduce variation into per-capita budget outcomes which does not reflect the behavior of policy makers, but instead reflects measurement error in the population data only revealed long after budgeting decisions have been made. In}
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, published by the Inter-University Consortium for Political and Social Research (ICPSR). Finally, 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

We first present results produced using the floating and fixed weight methods. We then present our preferred results produced using the simultaneous weight approach. Next, we explore fiscal outcomes other than taxes and total expenditures. Finally, we present several robustness checks.

4.1 Fixed and Floating Weight Results

Panel A of Figure 1 plots the per-capita tax collections of both Colorado and the synthetic Colorado. The synthetic collections are simply the weighted sum of collections from the states forming the synthetic Colorado. (We discuss the composition and observable attributes of the synthetic Colorado formed under the simultaneous weight method in section 4.2). 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, suggesting that TABOR had absolutely no effect on tax collections. The any case, our results are little changed if the revised data is used.
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.

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 tests; 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 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.

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. Indeed, as can be seen in Panel B, some placebo states demonstrate poor pre-treatment fits. We generally omit these states when performing the inference exercise. The criterion for exclusion is based on the mean square predicted error (MSPE) in the pre-TABOR period between actual and synthetic collections. We omit states that have an MSPE five times greater than that of Colorado.\textsuperscript{18} Panel C

\textsuperscript{18}Formally, $MSPE_i = (Z_i - Z_0W)'(Z_i - Z_0W)$. 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
displays the inference figure after the exclusions. Colorado remains firmly within the mass of placebo results.

An alternative inference method involves calculating the ratio of the mean post-treatment MSPE to the mean pre-treatment MSPE. The Colorado effect is ‘significant’ if its 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 D of Figure 1 displays the distribution of ratios. Colorado is in the lower end of placebo distribution providing no support for a TABOR effect.

Turning to expenditures, we first employ the fixed weight method – i.e. the weights which determine the composition of the synthetic Colorado are held fixed at the values produced for the tax outcome. We prioritize the tax outcome for two reasons. First, TABOR places the most constraint on this fiscal outcome and it is therefore the natural focal point of our study. Second, of all the outcomes we examine, taxes per capita achieves the best pre-treatment fit (both for Colorado and the placebo states).

Figure 2 displays the results. As expected, the pre-TABOR fit is not nearly as tight as the fit with taxes. Still, in the pre-TABOR period Colorado is centered around zero and well within the mass of placebo states. In the post-TABOR period there is no evidence of a decrease, if anything there may be an increase in the early 2000s. Due to the reduction in pre-treatment fit resulting from fixing the weights, the comparison of post/pre-TABOR MSPE ratios will be an important part of our interpretation of the inference (Panel B). Colorado is roughly in the middle of the placebo distribution providing no support for a TABOR effect.\(^\text{19}\)

The right-hand side of Appendix Figure A-2 displays the expenditure results when the criteria of five times the MSPE of Colorado.

\(^{19}\)Paralleling our treatment of holding the synthetic Colorado weights fixed, the synthetic weights for the placebo states are also held fixed to the values produced for the taxes per capita outcome.
synthetic weights are not held fixed, but instead are re-optimized – the floating weight method. The pre-treatment fit is again not nearly as tight as it is for taxes and there is again an absence of any evidence that TABOR restricted the size of government. If anything the estimates suggest that expenditures rose in the post-TABOR period. The post-TABOR treatment effect in near, or outside, the upper edge of the placebo distribution in a couple of years in the first-half of the 1990s and the first-half of the 2000s. 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.\textsuperscript{20}

These results are a bit of a puzzle when viewed in conjunction with the tax results on Figure 1: Taxes and expenditures are closely linked through the budget identity and should broadly move together over time. The tension may arise from the difference in the composition of the counterfactual Colorados used to identify the two treatment effects and illustrates why we do not prefer this approach. The other possible explanation 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.

4.2 Simultaneous Weight Results

Before presenting the simultaneous weight treatment effects, we discuss the synthetic Colorado produced by the 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. The next two columns display the mean values for the synthetic Colorado and the pool of 47

\textsuperscript{20}The expenditure results should be interpreted cautiously given that at several points in the pre-TABOR period the treatment effect is nearly outside the mass of placebo states (see panel B).
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 income per capita. It can be inferred that the synthetic matching procedure places less importance on these state attributes. As displayed on Table 2, which contains the $W$ weights, the synthetic Colorado is constructed from seven donor pool states. Four of these states, Arizona, Nevada and Utah are quite similar to Colorado in terms of region and the political composition of their state governments. These states comprise 82% of the synthetic Colorado. Hawaii, New Hampshire and New York make up the remainder.

Figure 3 displays the tax and expenditure results with the weights estimated simultaneously. The tax results—displayed on the left-hand side—are quite similar to those based on forming the weights solely on the basis of the tax outcome. In the post-TABOR period, tax collections in Colorado track collections in the synthetic Colorado extremely closely, suggesting TABOR did not lower tax receipts. In Panel C, we fail to reject the null that TABOR has no effect by a wide margin as Colorado is in the far left of the distribution. More formally, using a calculation similar to equation (10) to the test the hypothesis that TABOR failed to influence taxes, we obtain a p-value of 0.7 and fail to reject the null.

Given the magnitude of the TABOR rebates from 1997 through 2001—they ranged from $35 to $198 per person—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 tax policy for Colorado and its
synthetic counterpart. The tax policy change data is obtained from the NASBO reports – an independent data source from that used to estimate the synthetic cohort procedure. For the real Colorado, the change in tax policy liability includes the TABOR rebates (as measured in sources produced by the Colorado state government).21 Although there is some difference in the timing, both Colorado and the synthetic Colorado were reducing tax liability by roughly the same amount over this period. Strikingly, a taxpayer in either state saw his 2001 liability fall by somewhat over $200 relative to his 1995 liability. 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 nearly identical 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.

The expenditure results are on right-hand side of the Figure 3. As before, there is no evidence that TABOR lowered expenditures.

As noted above, debt financed capital expenditures may sever the connection between taxes and expenditures. Moreover, capital expenditures tend to be lumpy over time and may make it difficult for the synthetic cohort procedure to produce a synthetic pre-TABOR Colorado with expenditures closely matching those in the real Colorado. It is therefore interesting to examine operating expenditures – total expenditures minus capital outlays. These expenditures cannot generally be financed by debt. We therefore re-estimate the simultaneous procedure with the outcomes of total taxes and current operating expenditures. These are presented on Appendix Figure A-3. There is less evidence of a post-TABOR increase for operating expenditures and the synthetic Colorado also matches the real Colorado somewhat better in pre-TABOR period. Regardless, there is again no evidence that TABOR

---

21 The table displays the effect of all revenue legislation taken by the state 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 tax policy changes from 1996 plus the effect of the rebates.
constrained the size of government.

4.3 Other Outcomes

Figure 4 contains the results for three additional revenue outcomes: fees, total revenue and debt (all in per capita terms). In all cases, the synthetic Colorado is based on the weights produced by the simultaneous procedure run on taxes and expenditures (displayed on Figure 3). The placebo states are handled analogously. 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 nearly perfectly flat in the pre-period, indicating that the real and synthetic fees evolved in an identical manner. There is a noticeable upward trend starting in 2007, after the enactment of Referendum C. Referendum C lifted the TABOR revenue caps but left in place the requirement of voter approval for tax hikes. If policy makers wished to take advantage of the removal of the caps without going to voters for approval, fees were the logical budget lever to pull. However the accompanying histogram in the right of Panel A, which gives us some idea of significance over the entire period, suggests no effect. Furthermore, we acknowledge that the inference in the left of Panel A is problematic due to the divergence in levels between actual and synthetic fees in the pre-period. Thus, we interpret the fee results as being only suggestive in nature.

The results for total revenues, displayed in Panel B, exhibit a flat trend from 1977 until shortly after when total revenues in Colorado begin growing faster than 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 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 (see 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, may 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

The disagreement between our results and those of other publications is likely due to differing control groups. As previously stated, a key advantage of the synthetic control method is its data-driven choice of a counterfactual. Moreover, the inference method provides a formal approach for assessing the ability of the procedure to produce an appropriate counterfactual. Still, it remains an unverifiable assumption that the synthetic Colorado is an appropriate countefactual for Colorado in the period after TABOR. We therefore test the robustness of our conclusions to three alternative counterfactuals. We examine both per capita taxes and expenditures and re-estimate the weights in all cases using the simultaneous weight method. 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). These states share many of the geographic, economic, demographic and political characteristics of Colorado and intuition suggests they should provide a reasonable counterfactual for Colorado. Panel A of Figure 6 shows the results. 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 Colorados is fairly constant throughout the seven years leading up to TABOR and in 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. Indeed, 82% 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 will likely bias our estimates toward zero. 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. Similarly, policy makers may use their neighbors as a benchmark in deciding appropriate levels of taxation and spending. To address this concern, we run the synthetic control method after dropping all of the states bordering Colorado from the donor pool. As is visible in Panel C, there is a distinct downward drift in the tax collections of Colorado compared to the synthetic state. The descent begins around 1985 and ends around 1996 and consequently we cannot conclude that TABOR brought this about: 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. Panel D indicates that the synthetic state does a poor job of matching expenditures in the pre-TABOR period. However, there is no evidence that TABOR pushed down expenditures.

\footnote{Three of the five states in the primary synthetic Colorado are in the Mountain division–see Table 2.}

\footnote{Baicker (2005) and Case et al. (1993) provide evidence of policy spillovers between neighboring states.}
Third, if the states comprising the synthetic Colorado enact TELs during the sample period this may bias us against concluding that TABOR reduced the size of government. For instance, if a synthetic state passes an effective TEL in the post-TABOR period this will reduce synthetic tax collections and cause our tax treatment effect to increase spuriously. To guard against this possible source of bias, we restrict the sample to start in 1980 and drop states enacting a TEL between 1980 and the end of the sample. Thus, while TELs enacted before 1980 are in place in many of the donor pool states, the effect should be constant throughout the pre- and post-periods and should not confound the results. The synthetic state demonstrates good pre-treatment fit for taxes—see panel E—and displays no indication that TABOR was associated with a decline in taxes. In terms of expenditures, the pre-period fit is again poor and there is no evidence of a decline in expenditures associated with TABOR. (In several of the post-TABOR years the treatment effect is positive and outside the band of placebo states.)

A final robustness check concerns how we measure our fiscal outcomes. In all cases we have measured budget outcomes in per capita terms. Normalizing by personal income is a possible alternative. 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 7 presents tax and total expenditure results, both denominated by personal income, as a final robustness check. Weights from the
simultaneous method are used (displayed on Figure 3). The figures suggest that Colorado has unusual preferences. Relative to its synthetic state, with which it shares very similar per capita tax collections, it has unusually low taxes as a share of income. However, both taxes and expenditures display a relatively flat trend line over the entire sample period, providing no evidence of a TABOR effect.

5 Interpretation

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 its 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 signal 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 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.\textsuperscript{24} 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 choose to undermine the intent of their predecessors who passed TABOR. Amendment 23 increased permissible revenue growth and mandated increased in K-12 education spending.\textsuperscript{25} 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 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 suggest that such rules are unlikely to succeed in altering budget outcomes.

\textsuperscript{24}See Hill (2004); Eslinger (2005); Symanski (2010).
\textsuperscript{25}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.
6 Conclusion

In this paper we assess the efficacy of the nation’s most stringent tax and expenditure limitation in regards to its stated aim of restraining government growth. Using the synthetic control method we construct a counterfactual Colorado in order to estimate what fiscal outcomes would have been in Colorado in the absence of TABOR. The results suggest that TABOR had no effect on the size of government in the state of Colorado and that the resulting fiscal outcomes reflect preferences that would have been realized through some other channel in the absence of TABOR.
References


Figure 1: Total Taxes Per Capita

Panel A

Panel B

Note. All control states included.
Panel C

Note. Control states with pre-TABOR MSPE greater than 5 times Colorado’s pre-TABOR MSPE are excluded.

Panel D

Note. All control states included.
Figure 2: Total Expenditures Per Capita With Fixed Weights

Panel A

Panel B

Note. Panel A is produced with the fixed weight method using the weights from the tax outcome. Control states with pre-TABOR MSPE greater than 5 times Colorado’s pre-TABOR MSPE are excluded in Panel A.
Figure 3: Taxes and Expenditures With Simultaneous Weights

Panel A: Taxes and Expenditures in Synthetic vs. Real Colorado

Panel B: Inference for Taxes and Expenditures using Multiple Outcome Synthetic Controls

Panel C: MSPE Ratios of Taxes and Expenditures

Note. Control states with pre-TABOR MSPE greater than 5 times Colorado’s pre-TABOR MSPE are excluded in the figures in Panel B. All control states are included in the figures in Panel C.
Figure 4: Fees, Total Revenue and Debt

Panel A: Fees Per Capita

Panel B: Total Revenues Per Capita

Panel C: Debt 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.
Figure 5: Components of Spending

Panel A: K-12 Education Spending Per Capita

Panel B: Health and Hospitals Spending 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.
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 1980 or Later

Panel F: Expenditures, Computed Using Donor Pool that Excludes States that Enacted TELs in 1980 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 the sample is restricted to start in 1980 and any state enacting a TEL in 1980 or later is excluded from the donor pool.
Figure 7: Personal Income

Panel A: Taxes as Percent of Personal Income

Panel B: Expenditures as Percent of Personal Income

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.
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>A. Predictors of Budget Outcomes</th>
<th>Synthetic Colorado</th>
<th>Synthetic Colorado</th>
<th>All Donor Pool States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mountain Division</td>
<td>1.00</td>
<td>0.79</td>
<td>0.13</td>
</tr>
<tr>
<td>Republican Lower House</td>
<td>1.00</td>
<td>0.82</td>
<td>0.22</td>
</tr>
<tr>
<td>Republican Upper House</td>
<td>1.00</td>
<td>0.73</td>
<td>0.26</td>
</tr>
<tr>
<td>Democratic Governor</td>
<td>1.00</td>
<td>0.62</td>
<td>0.59</td>
</tr>
<tr>
<td>% of GSP from Manufacturing</td>
<td>13%</td>
<td>14%</td>
<td>20%</td>
</tr>
<tr>
<td>% of GSP from Finance</td>
<td>18%</td>
<td>21%</td>
<td>15%</td>
</tr>
<tr>
<td>% of GSP from Mining</td>
<td>4%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>State Unemployment Rate</td>
<td>6.1</td>
<td>6.1</td>
<td>6.6</td>
</tr>
<tr>
<td>Income per capita</td>
<td>24031</td>
<td>21390</td>
<td>21976</td>
</tr>
<tr>
<td>Pop. under 18</td>
<td>29.3%</td>
<td>32%</td>
<td>29.7%</td>
</tr>
<tr>
<td>Pop. over 64</td>
<td>8.7%</td>
<td>10%</td>
<td>11.2%</td>
</tr>
<tr>
<td>Education Profile:</td>
<td></td>
<td></td>
<td>0.0001</td>
</tr>
<tr>
<td>Pop. graduated high school</td>
<td>76.4%</td>
<td>73%</td>
<td>65.0%</td>
</tr>
<tr>
<td>Pop. with bachelor degree</td>
<td>21.9%</td>
<td>18%</td>
<td>15.2%</td>
</tr>
<tr>
<td>Population Growth (thousands)</td>
<td>49.1</td>
<td>55.3</td>
<td>47.8</td>
</tr>
</tbody>
</table>

| B. Per-Capita Tax Collections    |                    |                    |                      |
| Taxes per-capita 1980            | 1800               | 1821               | 1708                 |
| Taxes per-capita 1985            | 2116               | 2023               | 1971                 |
| Taxes per-capita 1990            | 2323               | 2302               | 2292                 |

Note. All cells display means. Pre-TABOR period refers to years 1977 through 1991. 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

<table>
<thead>
<tr>
<th>State</th>
<th>W Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ</td>
<td>0.42</td>
</tr>
<tr>
<td>HI</td>
<td>0.06</td>
</tr>
<tr>
<td>KS</td>
<td>0.03</td>
</tr>
<tr>
<td>NV</td>
<td>0.03</td>
</tr>
<tr>
<td>NH</td>
<td>0.04</td>
</tr>
<tr>
<td>NY</td>
<td>0.07</td>
</tr>
<tr>
<td>UT</td>
<td>0.34</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 3: Revenue Changes Due to Policy in Colorado and the Synthetic Colorado

<table>
<thead>
<tr>
<th>Year</th>
<th>Colorado</th>
<th>Synthetic Colorado</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>0</td>
<td>-49</td>
</tr>
<tr>
<td>1997</td>
<td>-34</td>
<td>-115</td>
</tr>
<tr>
<td>1998</td>
<td>-127</td>
<td>-156</td>
</tr>
<tr>
<td>1999</td>
<td>-153</td>
<td>-188</td>
</tr>
<tr>
<td>2000</td>
<td>-232</td>
<td>-193</td>
</tr>
<tr>
<td>2001</td>
<td>-250</td>
<td>-208</td>
</tr>
</tbody>
</table>

Note. Changes in real dollars per-capita. For Colorado these changes represent TABOR rebates plus cumulative net changes in revenue due to new policy in fiscal years 1996 through 2001. For the synthetic state they represent cumulative net changes in revenue due to new policy only.
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 (2008).
Figure A-2: Expenditures Per Capita With Synthetic States Weights Re-Estimated

Panel A

Note. Panel A is produced using the floating weight method – i.e. the weights are re-estimated based on the expenditure outcome. Control states with pre-TABOR MSPE greater than 5 times Colorado’s pre-TABOR MSPE are excluded in Panel A. All control states are included in Panel B.
Figure A-3: Taxes and Current Expenditures Per Capita With Simultaneous Weights

Panel A: Taxes

Panel B: Expenditures

Note. The weights are estimated using the simultaneous method with taxes and current expenditures as the two outcomes. 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.
### Table A1: TABOR Literature Review

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication or Organization</th>
<th>Conclusions in Regards to TABOR’s Effect on Taxes and Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Journal and Book Publications</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Billings, Stephen and Deborah Carroll (2012)</td>
<td>Growth and Change</td>
<td>TABOR restrained county and municipal government expenditures in the years immediately following its passage; see pg. 280-281. (Counties and municipalities where debruising 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 debruising initiatives were successful. In other words, being bound by TABOR had no effect on the creation of special district governments.)</td>
</tr>
<tr>
<td>Brown, Tom (2000)</td>
<td>Public Budgeting and Finance</td>
<td>TABOR constrained municipal budgets. The effects were not uniform across municipalities.</td>
</tr>
<tr>
<td>James, Franklin and Allan Wallis (2004)</td>
<td>Public Budgeting and Finance</td>
<td>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 &quot;de-Brucing,&quot; a process which allows them to retain excess tax revenues.</td>
</tr>
<tr>
<td>Kousser, Thad; McCubbins, Mathew; and Ellen Moule (2008)</td>
<td>Fiscal Challenges: An Interdisciplinary Approach to Budget Policy, eds. Garrett,</td>
<td>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.</td>
</tr>
<tr>
<td>Kousser, Thad; McCubbins, Mathew; and Kaj Rozga (2006)</td>
<td>State Politics &amp; Policy Quarterly</td>
<td>Colorado was the only state with a TEL that reduced per-capita spending; nevertheless, TABOR has been undermined with the 2005 passage of the &quot;timeout for TABOR&quot;.</td>
</tr>
<tr>
<td>Martell, Christine and Paul Teske (2007)</td>
<td>Public Administration Review</td>
<td>TABOR reduced the size of Colorado state government expenditures relative to the economy based on a comparison with other western states with a TEL.</td>
</tr>
<tr>
<td>New, Michael (2010)</td>
<td>State Politics &amp; Policy Quarterly</td>
<td>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.</td>
</tr>
</tbody>
</table>
### B. Selected Policy Pieces

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication or Organization</th>
<th>Conclusions in Regards to TABOR’s Effect on Taxes and Spending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atkins, Chris (2005)</td>
<td>Tax Foundation</td>
<td>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.</td>
</tr>
<tr>
<td>Bradley, David (2005)</td>
<td>Center on Budget and Policy Priorities</td>
<td>TABOR caused major state programs to be underfunded (e.g. higher education and children’s health care).</td>
</tr>
<tr>
<td>Citizens Budget Commission (2010)</td>
<td>Citizens Budget Commission</td>
<td>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 levels.</td>
</tr>
<tr>
<td>Greenwood, Daphne and Tom Brown (undated)</td>
<td>Center for Colorado Policy Studies</td>
<td>As municipal governments increasingly depend on sales tax revenues instead of property taxes, their revenue streams have become less stable; this situation is exacerbated by TABOR’s restrictions on spending growth.</td>
</tr>
<tr>
<td>Hedges, Carol (2003)</td>
<td>Bell Policy Center</td>
<td>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 levels.</td>
</tr>
<tr>
<td>Lav, Iris and Erica Williams (2010)</td>
<td>Center on Budget and Policy Priorities</td>
<td>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.</td>
</tr>
<tr>
<td>McGuire, Therese and Kim Reuben (2006)</td>
<td>Economic Policy Institute</td>
<td>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.</td>
</tr>
<tr>
<td>New, Michael and Stephen Slivinski (2005)</td>
<td>Cato Institute</td>
<td>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.</td>
</tr>
</tbody>
</table>
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.

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.