Testing Ricardian Equivalence With the Narrative Record on Tax Changes

ALFRED A. HAUG*

This version: April 2019

Abstract

The Ricardian equivalence hypothesis is tested empirically with a subcategory of the narrative measures of U.S. tax shocks developed by Romer and Romer (2010). The present value of tax increases motivated solely by concerns for improving the fiscal health of the government is used. These tax news represent a switch from debt to tax financing that should have no effects on real output and consumption. For the post-1982:IV period, fiscal anticipation plays an important role as many of the tax increases are implemented with substantial delays. Anticipated tax hikes increase economic activity in the delay period. Ricardian equivalence is rejected.

JEL Classification: C51, E62, H62

Keywords: Ricardian equivalence hypothesis; government budget deficits; narrative record; exogenous tax changes.

*Department of Economics, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand. *E-mail address*: alfred.haug@otago.ac.nz. The author thanks, without implicating, David H. Romer and participants at conferences for very helpful comments on an earlier version.

1. Introduction

An important benchmark for fiscal policy is the Ricardian equivalence hypothesis. It states that it does not matter for the economy whether government expenditures are financed by current taxation or instead by issuing government bonds (Barro, 1974). Ricardian households are only concerned about the present value of their intertemporal tax liabilities, which are in turn determined by the present value of the stream of government expenditures and currently outstanding government debt. The timing of taxes does not matter. Government bonds are simply seen as postponed taxes that will have to be paid for at some future date. Hence, a switch from tax to bond financing, or vice vera, has no effect on output or consumption. Barro (1974) argued that Ricardian consumers react, as long as the stream of government spending is kept fixed, to a tax cut financed by an increase in the government deficit by increasing savings by an equal amount. The tax cut leads to a dollar-for-dollar increase in bond holdings. Therefore, neither output, or output growth, nor interest rates change. The Ricardian equivalence hypothesis is based on several restrictive assumptions, such as lump-sum taxes, perfect capital markets where forward-looking households do not face liquidity constraints and can borrow at the same interest rate as the government, altruistic operative bequests that link generations, and no uncertainty about the future tax incidence.

A large number of theoretical and empirical papers have studied Ricardian equivalence, particularly during the period of high federal U.S. government budget deficit to GDP ratios in the 1980s. This literature has been surveyed by Seater (1993) and more recently by Ricciuti (2003). On a theoretical level, relaxing the very strict assumptions necessary for Ricardian equivalence can lead to government bonds having either positive or negative net wealth effects for households (Barro 1974, 1989). Under Ricardian equivalence, government bonds are not being considered net wealth. For example, Judd (1987) demonstrated this in a theoretical model with distortionary taxes, finite lives, and adjustment costs. Positive and negative wealth effects can cancel each other or be altogether negligible. Therefore, the fact that tax systems are generally not based on lump-sum taxation does not invalidate

Ricardian equivalence, as, for example, transportation costs do not invalidate the assumption of perfect competition in many applications. The issue cannot be settled on theoretical grounds.[1] However, the empirical evidence is not conclusive. The recent global financial crisis and the ensuing sovereign debt crises in Europe have sparked renewed research interest in fiscal policy issues. For example, Hayo and Neumeier (2017) study Ricardian equivalence with German population survey data, carried out in 2013, asking about changes in consumption and savings behaviour in response to large increases in public debt in the period from 2008 to 2012.

In this paper, we use the present value of narrative U.S. tax changes that were introduced in order to reduce inherited government budget deficits and not to finance new government spending. Therefore, such legislated tax news constitute a switch from bond financing of deficits to tax financing. If Ricardian equivalence is a good approximation, this switch should have no statistically significant effects on real GDP and real private consumption. In addition, another novel feature of our paper is that we study the role of exogenous deficit-driven tax changes that are implemented with a delay of several quarters, i.e., the role of fiscal foresight or anticipation for deficit-driven taxes.

Romer and Romer (2010) pioneered the method to constructed narrative U.S. tax shocks for the post-WWII period from a range of government documents.[2] However, they did not analyse separately the present-value time-series for "deficit-driven" tax revenue changes motivated by inherited budget deficits, introduced for reasons unrelated to current macroeconomic fluctuations or government spending. We use this subcategory of Romer and Romer's (2010) exogenous (as opposed to endogenous) deficit-driven tax *changes* in order to empirically test Ricardian equivalence. The goal is to shed new light on the controversial issue of whether the economy displays Ricardian equivalence features, which is of relevance in particular to orienting

[1] See Evans (1991). Furthermore, Evans *et al.* (2012) recently showed that rational expectations are not necessary for Ricardian equivalence to hold and a certain adaptive learning rule instead can produce equivalence. However, learning may happen only gradually and not uniformly across individuals, as demonstrated in a laboratory experiment by Meissner and Rostam-Afschar (2017).

[2] These tax schocks have been used by several others, such as Favero and Giavazzi (2012) and Mertens and Ravn (2012). Furthermore, the narrative approach was also employed by Ramey (2011) to study instead the fiscal multiplier effects of large U.S. government military spending shocks. See Ramey (2019) for a recent survey on fiscal multipliers.

theoretical models and for understanding some of the effects of fiscal consolidation.

Section 2 describes aspects of the data and theory relevant for the study and Section 3 presents results for the full post-WWII sample and for a sub-sample post 1982:IV. Accounting for a structural breaks, the sub-samples prior to 1982:IV have too few observations, due to further breaks, in order to allow for reliable inference. But the post-1982:IV sample delivers empirical results that show that deficit-driven tax increases have statically significant and positive effects on real output and real consumption during the implementation-lag period, i.e., before the announced tax increases take effect. The effects are negative after the tax changes are implemented. Ricardian equivalence is therefore rejected by the data. Moreover, the results support the presence of fiscal anticipation effects, contrary to Romer and Romer's (2010) analysis with aggregated exogenous tax shocks. The conclusion discusses further implication of the findings.

2. Exogenous Tax News Motivated by Deficit Reduction: The Data and Theory

2.1 Data Construction

Romer and Romer (2010) used in their empirical analysis several measures of exogenous tax news, all expressed as percentages of GDP: (a) The sum of the deficit-driven tax changes and the tax changes motivated by long-run economic growth concerns, as valued at the time of their implementation; (b) the deficit-driven tax changes, as valued at the time of their implementation; (c) the long-run tax changes, as valued at the time of their implementation; and (d) the present value of the stream of the sum of deficit-driven and long-run tax changes at the time when they were legislated, instead of values at the time of their implementation.[3] The data cover the period from the first quarter in 1945 to the fourth quarter in 2007.

This paper uses instead the present values of the deficit-driven tax changes at the time when they are legislated, as percent of GDP, which were not consider

[3]The three-year Treasury bond rate is used for calculating present values (Romer and Romer, 2010, p. 793, footnote 25). Also, the narrative events are described in detail in Romer and Romer (2009a).

separately by Romer and Romer (2010), nor subsequently by Favero and Giavazzi (2012) or Mertens and Ravn (2012), who also used Romer and Romer's data.[4] As pointed out by one of the anonymous referees, Romer and Romer's calculations are not based on actual tax changes but instead on government estimates of the quantitative effects of the tax changes at the time of legislation. For these calculations, it was generally assumed that the tax base remains the same after the legislated tax changes take effect.[5]

This paper additionally analyses the effects of anticipated and unanticipated exogenous deficit-driven tax changes. In contrast, previous studies on fiscal foresight included exogenous tax changes motivated by concerns over long-run economic growth.[6] Romer and Romer's (2009a, pp. 12-16) Table 1 records changes in tax liabilities by date and motivation for all 50 major U.S. federal tax changes in the postwar period. In particular, they included in the table the present values at the time of passage into law and in addition the values at the time of implementation of deficit-driven tax changes. It should be noted that the number of deficit-driven changes listed by implementation dates is much larger than for present values in their Table 1. Some legislation included both exogenous deficit-driven tax changes and increases in government spending. The two changes can be treated as different events when they are separated in time but not when they occur together at the time of passage into law. When the implementation date for deficit-driven tax changes was more than a year after passage, it was treated as a separate event and counted as an exogenous deficit-driven change and otherwise as an endogenous spending-driven tax change. The present value series excludes six such events. Also, the exogenous deficit-driven tax changes were all permanent tax changes, except for the change in

[4] On the other hand, Romer and Romer (2009b) studied exogenous tax cuts driven by long-run economic growth considerations in order to test the hypothesis that such tax reductions lead to decreases in future government spending.

[5] Mertens and Ravn (2013) presented econometric procedures for treating narrative income-tax changes as proxy variables for structural income-tax shocks in order to avoid potential estimation biases due to measurement error. We follow here instead Romer and Romer (2010), Favero and Giavazzi (2012), and Mertens and Ravn (2012), among others, and assume a one-to-one mapping between the narrative shocks and the true, but unknown, structural shocks. Also, our analysis relies on a very specific type of tax change, which may limit its general (external) validity.

[6] Favero and Givazzi (2012) and Mertens and Ravn (2012) utilized the sum of deficit-driven and long-run tax changes for their tax multiplier estimations.

the Federal-Aid Highway Act of 1959 that specified explicitly a temporary increase, scheduled to expire on 20 June 1961 (Romer and Romer, 2009a, p. 35).

Several tax laws implemented a specific tax change gradually over time at different dates. This leads to additional tax events when recording quarters with anticipated tax changes that have not yet been implemented. The full sample has 16 instances of anticipated exogenous deficit-driven tax changes that are implemented with delays and six unanticipated exogenous deficit-driven tax changes that are implemented without delays and are hence treated as surprise tax shocks. Following Mertens and Ravn's (2012, pp. 152-153) method of constructing anticipated tax-change variables, we record for each anticipated deficit-driven tax shock its remaining anticipation horizon in a given quarter. Next, we measure the sum of all anticipated deficit-driven tax changes known at date t to be implemented at date $t+i$, denoted $\tau_{t,i}^{a,\,def}$, so that we construct a different anticipation variable for every anticipation horizon that records the value of all anticipated tax changes at time t that occur in each of the i= 1, 2, 3, 4, ..., and M quarters in the future. The largest implementation lag is M, so that there are M separate anticipation variables for each time period t. However, for reasons of parsimony, Mertens and Ravn (2012) restrict the number of anticipation variables to six, the median implementation lag in their sample.

Most of the exogenous deficit-driven tax changes were motivated by concerns over the long-run solvency of the U.S. Social Security system. For example, the Social Security Amendments of 1977 and 1983 were major tax increases that did not simultaneously increase benefits. The largest deficit-driven tax increases not related to Social Security were those in the Tax Equity and Fiscal Responsibility Tax Act of 1982 and the Omnibus Budget Reconciliation Acts of 1990 and 1993. There are ten (nine) quarters with exogenous deficit-driven tax changes passed into law in the sample period from 1945:I to 2007:IV (1950:I to 2007:IV). This is a considerably larger number than the four quarterly non-zero observations for the war dummy variable in Ramey (2011). Figure 1 depicts the present values of deficit-driven tax changes as percent of GDP. The values range from 0.023% in 1997:III to 1.153% in 1977:IV. In comparison, the non-zero defense build-up dates in Ramey (2011, Table II, pp. 26-27) represent the following shares of GDP: 63.06% in 1950:III, 0.33% in 1965:I,

6.36% in 1980:I, and 0.94% in 2001:III. Two of these events show much larger portions of GDP than values in our sample, though not the remaining ones. In addition, none of the deficit-driven dates overlap with the dates for the present values of long-run exogenous tax changes of Romer and Romer (2010).

Following Romer and Romer (2010), regressions start in 1950:I in order to allow for sufficient lags drawing on observations from earlier quarters. Single-equation regressions involve the narrative tax shocks and assess their effects on real GDP growth rates, taken from Romer and Romer (2010).[7] The vector autoregressions (VARs) use endogenous variables, retrieved from Favero and Giavazzi (2012):[8] the log of real GDP per capita, the log of real federal primary (excluding net interest payments) government expenditures per capita, the log of real federal government tax receipts per capita, inflation measured with the log difference of the GDP deflator, and an interest rate that measures the average nominal cost of financing government debt.[9] Blanchard and Perotti (2002) used instead a VAR with only the first three variables, however, it has become standard to use such a five-variable VAR as a baseline model for fiscal multiplier analysis following their approach (Ramey, 2019). For one specification of the VAR, we add the log of real personal consumption expenditure per capita as an endogenous variable. Regressions were carried out with the software packages EViews, GAUSS, and WinRATS.

2.2 A Brief Outline of Theory in Relation to the Data

When Ricardian equivalence holds, the permanent income hypothesis implies that households incorporate the intertemporal government budget constraint into their permanent income. The present values of the exogenous deficit-driven tax shocks have been selected so that there are no associated changes to government spending announced, i.e., there are no government spending shocks at the same time.[10] This

[7] See https://www.aeaweb.org/articles.php?doi=10.1257/aer.100.3, where their RATS software code is also available.

[8] https://www.aeaweb.org/articles?id=10.1257/pol.4.2.69

[9] Using instead the effective Federal Funds rate (https://fred.stlouisfed.org/series/FEDFUNDS) as the nominal interest rate has only minimal effects on our results. It is available from July 1954 onwards.

[10] In a regression context this means that such narrative variables are, as long as orthogonality holds, uncorrelated with other included and other omitted regression variables, and their regression

is exactly the scenario needed for testing the Ricardian equivalence hypothesis, i.e., keeping the expected stream of government spending fixed when a tax news shock hits.

Would Romer and Romer's (2010) tax cuts motivated by concerns about long-run growth be suitable for testing Ricardian equivalence? The long-run tax-change variable was constructed so that it is orthogonal to the current business cycle and current government spending. However, the stated intention of these tax shocks was explicitly to either decrease the size of the government (e.g., the 1981 Reagan tax cuts), or to spur long-run productivity growth and increase efficiency.[11] This subcomponent of tax changes, which stimulates long-run growth, is therefore not suitable for testing Ricardian equivalence because there are associated changes in *future* government spending or in *future* income that both affect permanent income.

In order to illustrate the Ricardian case, it is useful to look at a simplified version of the intertemporal government budget constraint (see, e.g., Hakkio and Rush, 1991, and Haug, 1996):[12]

$$\sum_{s=1}^{\infty}(1+r)^{-s}RE_{t+s} = B_t + \sum_{s=1}^{\infty}(1+r)^{-s}GE_{t+s} - \lim_{s\to\infty}(1+r)^{-s}B_{t+s},$$

where RE_t is real tax revenue in period t, including revenue from monetizing deficits by printing money (which is treated as an inflation tax). B_t denotes real government debt, r is the real interest rate, assumed constant, and GE_t stands for real government expenditures on purchases of goods and services plus transfer payments. It is assumed that the government does not follow a Ponzi scheme so that intertemporal government budget balance holds and the limit term in the above equation goes to zero.

Ricardian households do not change their consumption plans as long as the present value of expected future tax revenue, which is equal to current bonds and expected future government spending, does not change. The future path of govern-

coefficient estimates are unbiased. The only effect of omitted variables is to increase the residual variance. It is therefore possible, in principle, to analyse the effects of exogenous tax changes on economic activity, such as real GDP, without specifying an economic model that includes other fiscal and monetary policy variables.

[11] Whether deficits (and government debt) in general affect long-term economic growth is a related, but different question for which the empirical evidence has not been conclusive to date.

[12] It is straightforward to define this intertemporal government budget constraint instead in terms of ratios to GDP in order to illustrate the effects (see, e.g., Hakkio and Rush, 1991).

ment debt is irrelevant because all debt has to be ultimately financed by taxes. The above deficit-driven tax news fit this scenario under our null hypothesis of Ricardian equivalence, because the tax increases are explicitly targeting a reduction in inherited deficits and therefore lead to less national debt. Less national debt means an offsetting reduction (in present value terms) in implied future taxes. Hence, a deficit reduction now is a switch from future to current taxation, while the stream of government spending is expressly kept fixed. In other words, the two different time paths of taxes (before and after the deficit-driven tax news) have the same present values under the null hypothesis.

When Ricardian equivalence does not hold, the permanent income hypothesis predicts that consumers react to news about current and future tax increases (decreases) by immediately reducing (increasing) consumption. This initial reaction could then possibly lead to additional dynamics. However, there should be no further reaction at the time when tax changes actually take effect, if they are not implemented at the time of being passed into law. The correct approach under the permanent income hypothesis for dealing with tax news is therefore to use the present value of tax news at the time when they are announced and to relate output movements to current values of such a measure of news about tax changes. This scenario holds with lump-sum taxes. When distortionary taxes are considered and Ricardian equivalence does not hold, intertemporal substitution predicts an increase in consumption before a tax increase takes hold and a reduction in consumption once the increased tax takes effect. In other words, spending is moved forward in time to take advantage of lower taxes before the tax increase is implemented. This will lead to an increase in real consumption and output in the period between the announcement and implementation of a tax increase, at the expense of lower real consumption and output after the implementation date. Leeper *et al.* (2008) referred to this as fiscal foresight or anticipation. In this case, the model needs to account for both the date when the tax change is announced (legislated) and the dates when it is implemented.

The permanent income hypothesis assumes that consumers are not liquidity constrained and are able to borrow and smooth consumption intertemporally without restrictions. On the other hand, Keynesian models assume instead that consumption

is determined by current disposable income because of borrowing constraints, so that Ricardian equivalence cannot hold. This implies that consumers react to tax changes when they affect their disposable income, which is the quarter when they are implemented and not when they are announced or during the implementation lag period.

3. Ricardian Equivalence Test Results: OLS and VAR

3.1 Full Sample Period: OLS Without and With Controls for Output Growth

The first step in the analysis of Romer and Romer (2010) is to study the effect of an exogenous tax change on real output. Instead of using total exogenous tax changes, we select the present value of deficit-driven tax changes:

$$\Delta Y_t = \alpha + \sum_{i=0}^{12} \beta_i \Delta T_{t-i} + \varepsilon_t, \tag{1}$$

where ΔY_t is the log change (growth rate) of real GDP, ΔT_t is the present value of the stream of deficit-driven tax changes legislated at time t, expressed as percent of GDP, and ε_t is a mean-zero Gaussian error that is orthogonal to ΔT_{t-i}. The advantage of the specification in equation (1) is that the β_i coefficients represent (impact) tax multipliers in percentage terms. The sum $\sum_{i=0}^{j} \beta_i$ measures the cumulative fiscal multiplier from period 0 to period j.

We can formulate the following hypotheses:

Hypothesis 1. Ricardian equivalence predicts that the effects from a deficit-driven tax increase on real GDP should be statistically insignificant, so that all $\beta_i = 0$.

Hypothesis 2. The permanent income hypothesis predicts for tax increases a negative effect on real output and consumption under Ricardian non-equivalence, as expected after-tax real income falls and government bonds are treated as net wealth so that overall household wealth falls. This causes additional dynamic effects that lower real GDP so that $\beta_i < 0$ for some or all i. However, if a tax increase is implemented with a delay, intertemporal substitution predicts, in the presence of distortionary taxes and Ricardian non-equivalence, an increase in consumption (or investment, depending on the type of tax) during the period of delay before the implementation takes hold.

Hence, $\beta_i > 0$ for periods i till the implementation date. Once the tax increase takes effect, real output falls and $\beta_i < 0$ for periods i after the implementation date.

Hypothesis 3. A Keynesian model suggests that a tax change affects real GDP and consumption only once the tax change takes effect, i.e., has been implemented.

Figure 2 presents the cumulative effect of a 1% of GDP increase in deficit-driven tax revenue (solid line) on real GDP, estimated with ordinary least squares (OLS). The effect on real GDP is negative in the first period and positive thereafter, reaching a peak of 1.98% five periods after the impact of the tax shock and then declining. Are the effects of deficit driven tax increases statistically significantly different from zero? The answer depends on the width of the confidence bands used in Figure 2. The usual two standard-error confidence band, which is for normally distributed coefficient estimates approximately a 95% band, indicates no statistical significance at any horizon in Figure 2. Using instead the one standard-error confidence band (an approximate 68% confidence band), gives a different picture. Now, the responses of real GDP are statistically significantly different from zero at horizons three to eight after the tax shock. However, the responses are positive so that the tax increase causes an increase in real GDP at these horizons, which is the opposite direction of what the non-Ricardian theories predict.

The next step is to include in equation (1) the lags of GDP growth in order to check the robustness of the above results:

$$\Delta Y_t = \alpha + \sum_{i=0}^{12} \beta_i \Delta T_{t-i} + + \sum_{i=1}^{11} \gamma_j \Delta Y_{t-j} + \varepsilon_t. \tag{2}$$

The lags of output growth control for the dynamics of GDP and a "multitude of other influences" (Romer and Romer, 2010, p. 781). Figure 3 reports the effects of a 1% of GDP increase in deficit-driven taxes on real GDP when lagged output growth is controlled for in the regression. Overall the effects are similar in magnitude to those without controls, though the peak effect is lower. The one and two standard-error confidence bands are calculated from bootstraps with 10,000 replications. The effect in the first period is again negative but remains statistically insignificant for the 95% confidence band. The 68% confidence band indicates statistical significance at the

first horizon, although just barely so. The only other horizon where there is possibly statistical significance is the fourth one, but again it is very much a borderline case.

3.2 Testing for Structural Change

The empirical evidence over the full sample period would seem to favour the Ricardian equivalence hypothesis. But, the crucial question is whether the regressions are structurally stable. In the postwar period, different fiscal and monetary policies have been in place. It is therefore possible that structural breaks may lead to misleading statistical inferences. The structural change tests of Bai and Perron (1998, 2003) allow estimating the number of breaks and the break dates consistently when break dates are not known. They also provided tests for assessing the statistical significance of breaks. Bai and Perron (2003) developed an efficient algorithm to obtain global minima for the sum of squared residuals. Their econometric framework accounts for possible heteroscedasticity and autocorrelation in the residuals.[13,14]

As recommended by Bai and Perron (2003, p.16), we first apply the double maximum test UDmax that is based on sequential F-type tests of the null hypothesis of no break against the alternative hypothesis of one break at an unknown date. This test allows establishing whether there are any breaks at all in the sample. If it indicates breaks, we apply next Bai and Perron's sequential sup-F test of the null hypothesis of l breaks against the alternative hypothesis of $l+1$ breaks in order to determine the number of breaks and the unknown break dates. The testing procedures require to trim the endpoints of the sample. We choose alternative sample trimming parameters ε equal to 0.05, 0.10, and 0.15 in order to explore the sensitivity of the results. We calculated a test statistic value of 15.72 for the UDmax test and a corresponding simulated 1% critical value of 13.74. The null hypothesis of stability is comfortably rejected. The sequential sup-F test indicates five statistically significant breaks, with new regimes starting in 1953:I, 1960:II, 1968:II, 1979:3 and 1982:IV. There is some sensitivity to the trimming of the endpoints imposed, and ε=0.10 or

[13] A Breusch-Godfrey LM test indicates serial correlation in the residuals of equation (2): ($p = 0.03$ (one lag) and 0.002 (four lags). This invalidates cumulative sum (CUSUM) type tests for breaks.

[14] We use a data-dependent quadratic spectral kernel with automatic bandwidth selection and AR(1) pre-whitening to estimate autocorrelation and heteroskedasticity robust covariance matrices.

0.15 instead of 0.05 leads to the same number of breaks but slightly different break dates. For the last break, the dates are in the period from 1982:II to 1983:I. We choose 1982:IV for the start of the new regime, which is when monetary targeting was abandoned by the U.S. Federal Reserve and is consistent with other studies (e.g., Chang and Kwak, 2017). The effect of fiscal policies (the fiscal multipliers) depend on the stance of monetary policy and it is therefore likely that a change in the monetary regime also changes the fiscal regime.

3.3 The Post-1982:IV Period: OLS With Controls for Output Growth

The sub-samples identified with the structural change tests in the previous section prior to 1982:IV are too small and have too few deficit-driven tax changes for reliable analysis. Therefore, we focus on the post-1982:IV period. This period has 101 quarterly observations and includes six of the ten tax changes in present value terms.[15] Figure 4 shows for equation (2) the cumulative effect for the post-1982:IV period of a 1% of GDP increase in deficit-driven tax revenue on GDP. The multiplier is negative in period 0 and period 1 but statistically insignificant. From period 2 onwards it is positive, reaching the largest value of 2.81 in period 12. It is statistically significant only at the 68% level and not at the 95% level from period 2 to 12. However, we have not yet considered a VAR specification and fiscal foresight in deriving this result.

3.4 The Post-1982:IV Period: Unanticipated and Anticipated Tax Shocks in a VAR

An alternative model specification to single equations is a VAR. A single equation approach is based on limited information. A VAR allows for rich dynamics and endogeneity of the variables included in the VAR. We expand on the limited dynamic structure captured in equation (2), with only lags on output growth, by including endogenous variables in the VAR together with exogenous unanticipated and anticipated tax shocks for modelling fiscal foresight.[16] We follow Favero and Giavazzi

[15] The period from 1968:II to 1979:II has the largest number of observations among the other sub-samples, which is 48, but only one deficit-driven event occurs in this period.

[16] A VAR for the model below with only the present values of exogenous deficit-driven tax shocks, and no other tax shocks, leads to impulse responses very similar to those in Figure 4 that were based on equation (2).

(2012, equation system (11), p. 89)[17] and use the narrative tax shocks directly for identifying structural shocks, instead of imposing externally calculated elasticities and short-run exclusion restrictions for the effects of structural shocks as in Blanchard and Perotti (2002). This requires that the narrative shocks are orthogonal to the variables included in the VAR.[18]

The VAR is specified as

$$
\begin{aligned}
R_t &= \alpha_1 + \mathbf{a}_1\mathbf{y}_{2,t-1} + \ldots + \mathbf{a}_p\mathbf{y}_{2,t-p} + \sum_{i=1}^{p} a_i R_{t-i} + \delta_{11}\tau_t^{un,\,def} + \theta_{11}\tau_{t,0}^{an,\,def} + e_{1,t} \\
\mathbf{y}_{2,t} &= \boldsymbol{\alpha_2} + \mathbf{B}_1\mathbf{y}_{2,t-1} + \ldots + \mathbf{B}_p\mathbf{y}_{2,t-p} + \sum_{i=1}^{p} \boldsymbol{b_i} R_{t-i} + \boldsymbol{\delta_2}\tau_t^{un,\,def} + \boldsymbol{\theta_2}\tau_{t,0}^{an,\,def} \\
&\quad + \sum_{i=1}^{12} \boldsymbol{\lambda_i}\tau_{t,i}^{an,\,def} + \mathbf{e}_{2,t},
\end{aligned}
$$

with $\mathbf{y}_t' = [R_t \quad \mathbf{y}_{2,t}']$ a $(n+1)\times 1$ vector of five variables at time t; α_1 is a constant and $\boldsymbol{\alpha_2}$ is a $n\times 1$ vector of constants; $\mathbf{a}_\ell$ is a $n\times 1$ coefficient vector and $\mathbf{B}_\ell$ is a $n\times n$ matrix of coefficients for $\ell = 1,\ldots,p$; a_i and $\boldsymbol{b_i}$ represent coefficients, as do δ_{11}, θ_{11}, $\boldsymbol{\delta_2}$, and $\boldsymbol{\lambda_i}$; and $e_{1,t}$ and $\mathbf{e}_{2,t}$ are mean-zero Gaussian reduced-form errors. Our baseline model consists of the following variables $\mathbf{y}_t$, as described in the data section: ln real federal government tax receipts per capita, denoted R_t; ln real GDP per capita; ln real federal primary government expenditures per capita; inflation; and the interest rate. The tax shocks $\tau_t^{un,\,def}$ are unanticipated exogenous deficit-driven tax changes that are legislated at time t and implemented in the same quarter t. The tax shocks $\tau_{t,0}^{an,\,def}$ are anticipated tax changes that have been passed into law in a previous quarter and are implemented in period t. The tax shocks $\tau_{t,i}^{an,\,def}$ are anticipated tax changes at time t that will be implemented in i quarters from period t, at time $t+i$. All tax shocks are calculated as percentages of GDP. Mertens and Ravn (2012) restrict the maximum number of periods i to equal the median implementation

[17] The omission of lags on the tax revenue variable in their second equation of the VAR system is a typographical error. Their online EViews code includes lags on tax revenue.

[18] It is difficult to establish when a tax change becomes anticipated so that it is no longer a shock (or news). Is it when a tax change is put forward for the first time by a politician, or when it is debated for the first time by Congress, or instead when it is passed into law? Also, do all economic agents get the information at the same time? In the end, the question is whether the constructed narrative tax shocks are predictable from past values of the variables (i.e., from the information set included) in the VAR.

lag in their sample. The average delay in the post-1982:IV sample is 8.5 quarters and the median delay is seven quarters. However, to be consistent with Romer and Romer (2010), we allow for 12 quarters.

In order to select the appropriate lag length p for the above VAR, we use a sequential likelihood ratio test with a small sample adjustment and a 5% level of significance (Sims, 1980; Kilian and Lütkepohl, 2017). We allow for a maximum of eight lags and find that three lags are statistically significant. This specification also produces residuals that are serially uncorrelated, based on a Lagrange multiplier test with up to fourth order serial correlation. Furthermore, no roots of the characteristic polynomial lie outside the unit circle so that this VAR specification satisfies the stability condition. In addition, we test whether the constructed narrative tax shocks are predictable from past values of the endogenous variables in the VAR. For the post-1982:IV sample, we regress by ordinary least squares the unanticipated tax shocks, $\tau_t^{un,\,def}$, on the information set used in the VAR with three lags: $\mathbf{y}_{t-1}$, $\mathbf{y}_{t-2}$, and $\mathbf{y}_{t-3}$. An F-statistic for the null hypothesis of no statistically significant relationship with the regressors (non-predictability) has a p-value of 0.70 so that the null is clearly not rejected. Hence, unanticipated tax shocks are not predictable. We repeat the same regression but with anticipated tax shocks at the time of their implementation, $\tau_{t,0}^{an,\,def}$, as the dependent variable. The p-value of the F-statistic is 0.52 so that we confirm non-predictability for this type of tax changes as well.[19]

Figure 5a shows the impact of a 1% of GDP positive revenue tax shock on real GDP ($\tau_t^{un,\,def}$), period-by-period, after the initial shock. The tax shock considered is an unanticipated increase in the tax liability that is implemented in the same quarter in which it is legislated. The peak effect is a decrease in real GDP of 0.68% at horizon one. The magnitude of the impact diminishes afterwards and becomes positive from quarter nine onwards. However, the effects are statistically significant only up to

[19] An ordered probit regression for the predicability of the dates of anticipated tax changes produces a test statistic with a p-value of 0.27 for the null hypothesis of non-predictability. A probit regression for unanticipated tax change dates does not produce reliable results due to a lack of non-zero observations. Further, the present-value series for the deficit-driven tax changes for the full sample period produces a p-value of 0.17 for the least squares regression and of 0.25 for the probit regression for dates of changes. Hence, non-predictability cannot be rejected either.

quarter four at the 95% confidence level and up to quarter six at the 68% level.[20] Figure 5b depicts instead the cumulative effects of this unanticipated tax increase. The cumulative multiplier of a 1% of GDP tax increase peaks at quarter eight with a value of -4.22%, i.e., real GDP falls by 4.22%. The effects are statistically significant at all quarters, except for the last one for the 95% confidence level. The signs of the effects are consistent with the permanent income hypothesis and also with Keynesian theory because such unanticipated tax code changes are implemented without delay and affect disposable income and permanent income at the same time. Furthermore, the magnitudes of this tax multiplier are broadly consistent with what other studies have reported.[21]

Figure 5c illustrates the impact of a positive anticipated tax revenue shock ($\tau_{t,0}^{an,\,def}$), period-by-period, that is equal to 1% of GDP. The impact on real GDP is largest in quarter one with a 0.39% increase in real GDP and falls thereafter. It becomes negative from quarter nine onwards, falling to -0.15% in quarter 12. These effects are statistically significantly different from zero up to quarter five for the 95% confidence interval and up to quarter six for the 68% interval. This result is consistent with the permanent income hypothesis that predicts an increase in real GDP in periods before the tax increase is implemented, followed by a decrease after the implementation date. The average implementation lag of tax shocks is 8.5 quarters in our post-1982:IV sample. On the other hand, this result is not consistent with household who consume based on disposable income only, because there should be no statistically significant effects on real GDP during the delay period, i.e., the first 8.5 quarters, and statistically negative effects should be present from the implementation date onwards, i.e., from nine quarters after the shock onwards. Figure 5d reports the cumulative effects of this anticipated positive tax shock. The multiplier reaches a maximum of 2.40% in quarter eight and falls afterwards, reaching 1.98% in quarter 12.

[20] All confidence intervals are based on 10,000 bootstrap replications.

[21] See Ramey (2019). She also described various alternative ways in which tax multipliers have been calculated in the literature. To make our results directly comparable to those of Romer and Romer (2010), Favero and Giavazzi (2012), and Mertens and Ravn (2012) we use their definition of the tax multiplier and not Ramey's (2019).

3.5 Robustness of the VAR Results for the Post-1982:IV Period

We explore the sensitivity of our results to various changes. Our analysis above is robust to changing the delay period to seven, eight or nine quarters instead of 12 for $\tau_{t,i}^{an,\,def}$ in the VAR. Generally, varying the maximum anticipation horizon leads to similar impulse response functions, which is what Mertens and Ravn (2012, pp. 158-159) found as well for their (different) tax shocks. Further, changing the exact break date, 1982:IV, by a quarter or two does not alter the qualitative results and leads to only very small quantitative variations. Next, we consider the sensitivity of the above results to large and small tax changes. For this purpose, we drop, in turn, the largest and smallest values for unanticipated tax changes, $\tau_t^{un,\,def}$, and anticipated tax changes, $\tau_{t,0}^{an,\,def}$, in the VAR: 1.025% of GDP in 1993:III and 0.078% in 1993:IV for the former, and 0.780% in 1983:I and 0.006% in 2002:I for the latter. The fiscal multipliers for a 1% of GDP change in taxes stay petty much the same, so that results are not driven by single events.

In the theoretical discussions on Ricardian equivalence, household consumption plays a crucial role. In Figure 6a, we study the impact of an unanticipated tax change on real consumption expenditure instead of on real GDP. We augment the above VAR with the natural logarithm of real personal consumption expenditure per capita.[22] Figure 6b demonstrates the impact on consumption of an anticipated tax shock instead. A comparison with the effects on GDP in Figures 5a and 5c reveals that the multipliers are smaller, though the impulse response functions are very similar otherwise. In Figure 6a, the peak effect is a decrease in real consumption of 0.19% at horizon two, instead of a decrease of 0.68% at horizon one for real GDP above. The magnitude of the impact diminishes thereafter and becomes positive from quarter 10 onwards. The effects are statistically significant up to quarter five at the 95% confidence level and up to quarter six at the 68% level. In Figure 6b, the impact on real consumption is largest in quarter two with an increase of 0.28%, compared to a 0.39% increase in real GDP at quarter one in Figure 5c. The impact falls after quarter two and becomes negative from quarter ten, falling to -0.08% in quarter 12. These

[22] The chosen lag length stays the same, the VAR residuals show no significant autocorrelation, and the system is stable.

impacts are statistically significantly different from zero up to quarter five for the 95% confidence interval and up to quarter seven for the 68% interval. In regards to the permanent income hypothesis, these results confirm the findings with real GDP. The Ricardian hypothesis is rejected by the data. The empirical results are again consistent with intertemporal substitution due to fiscal anticipation or foresight as argued by Leeper *et al.* (2008).

4. Conclusion

This paper contributes to the empirical literature on fiscal policy by using a part of a narrative measures developed by Romer and Romer (2010) not used previously. The paper provides new empirical evidence on the validity of the Ricardian equivalence hypothesis. Ricardian equivalence may be a good approximation to reality and the fact that taxes are generally not lump-sum does not necessarily invalidate it (Judd, 1987). The present values of U.S. exogenous deficit-driven tax changes is employed in order to shed new light on an unresolved empirical question.

The empirical results do not support the Ricardian equivalence hypothesis because the effects of a tax increase, keeping government spending fixed, have a statistically significant influence on real GDP and consumption. While Ricardian equivalence seemed to be supported at first sight over the full post-WWII sample period, this result does not hold up when the sample is split in 1982:IV. The post-1982:IV sample shows that increasing taxes solely motivated by concerns over the health of the fiscal position of the government has a statistically significant positive effect on real GDP and on consumption for anticipated tax increases. These positive effects occurs during the implementation-lag period and are predicted by the theory of fiscal anticipation of Leeper *et al.* (2008). Intertemporal substitution leads to a fiscal stimulus ahead of the implementation date for a tax increase but it comes at the expense of lower economic activity afterwards, once the tax increase takes effect. On the other hand, unanticipated tax changes that are implemented without delay, have statistically significant negative effects on real GDP and consumption. Overall, these results are consistent with the recent findings of Hayo and Neumeier (2017), using

survey data, and Meissner and Rostam-Afschar (2017), using a laboratory experiment and lump-sum taxation. Both studies revealed significant deviations from Ricardian equivalence. Furthermore, the results in this article are relevant for some aspects of recent fiscal consolidation policies.

References

Bai, J. and Perron, P. (2003). 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics,* Vol. 18, pp. 1-22.

Bai, J. and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes', *Econometrica,* Vol. 66, pp. 47-78.

Barro, R. J. (1989). 'The Ricardian approach to budget deficits', *Journal of Economic Perspectives,* Vol. 3, pp. 37-54.

Barro, R. J. (1974). 'Are government bonds net wealth?', *Journal of Political Economy,* Vol. 82, pp. 1095-117.

Blanchard, O. and Perotti, R. (2002). 'An empirical characterization of the dynamic effects of changes in government spending and taxes on output', *Quarterly Journal of Economics,* Vol. 117, pp. 1329-1368.

Chang, Y. and Kwak, B. (2017). *U.S.Monetary-Fiscal Regime Changes in the Presence of Endogenous Feedback in Policy Rules,* CAEPR Working Paper No. 2017-016. Available at: http://dx.doi.org/10.2139/ssrn.3080558 (last accessed: 23 March 2019).

Cloyne, J. (2013). 'Discretionary tax changes and the macroeconomy: new narrative evidence from the United Kingdom', *American Economic Review,* Vol. 103, pp. 1507Ű1528.

Evans, G. W., Honkapohja, S. and Mitra, K. (2012). 'Does Ricardian equivalence hold when expectations are not Rational?', *Journal of Money, Credit, and Banking,* Vol. 44, pp. 1259-1283.

Evans, P. (1991). 'Is Ricardian equivalence a good approximation?', *Economic Inquiry,* Vol. 29, pp. 626Ű644.

Favero, C. and Giavazzi, F. (2012). 'Measuring tax multipliers: the narrative method in fiscal VARs', *American Economic Journal: Economic Policy,* Vol. 4, pp. 69-94.

Hakkio, C. S. and Rush, M. (1991). 'Is the budget deficit 'too large'?', *Economic Inquiry,* Vol. 29, pp. 429-445.

Haug, A. A. (1996). 'Blanchard's model of consumption: an empirical study', *Journal of Business and Economic Statistics,* Vol. 14, pp. 169-177.

Hayo, B. and Neumeier, F. (2017). 'The (in)validity of the Ricardian equivalence theorem Ű findings from a representative German population survey', *Journal of Macroeconomics*, Vol. 51, pp. 162-174.

Judd, K. (1987). 'Debt and distortionary taxation in a simple perfect foresight model', *Journal of Monetary Economics,* Vol. 20, pp. 51-72.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis.* Cambridge University Press, Cambridge, UK.

Leeper, E. M., Walker, T. B. and Yang, S.-C. S. (2008). *Fiscal Foresight: Analytics and Econometrics,* NBER Working Paper No. 14028.

Meissner, T. and Rostam-Afschar, D. (2017). 'Learning Ricardian Eequivalence', *Journal of Economic Dynamics and Control,* Vol. 82, pp. 273-288.

Mertens, K. and Ravn, M. O. (2013). 'The dynamic effects of personal and corporate income tax changes in the United States', *American Economics Review,* Vol. 103, pp. 1212-1247.

Mertens, K. and Ravn, M. O. (2012). 'Empirical evidence on the aggregate effects of anticipated and unanticipated U.S. tax policy shocks', *American Economic Journal: Economic Policy,* Vol. 4, pp. 145-181.

Ramey, V. A. (2019).*Ten years after the financial crisis: what have we learned from the renaissance in fiscal research?,* NBER Working Paper No. 25531.

Ramey, V. A. (2011). 'Identifying government spending shocks: it's all in the timing', *Quarterly Journal of Economics,* Vol. 126, pp. 1-50.

Ricciuti, R. (2003). 'Assessing Ricardian equivalence', *Journal of Economic Surveys,* Vol. 17, pp. 55-78.

Romer, C. D. and Romer D. H. (2010). 'The macroeconomic effects of tax changes: estimates based on a new measure of fiscal policy shocks', *American Economic Review,* Vol. 100, pp. 763-801.

Romer, C. D. and Romer, D. H. (2009a). 'A narrative analysis of postwar tax changes'. Available at: https://www.aeaweb.org//articles.php?doi=10.1257/aer.100.3 (last accessed: 10 November 2017).

Romer, C. D. and Romer, D. H. (2009b). 'Do tax cuts starve the beast? The effects of tax changes on government spending', *Brookings Papers on Economic Activity,* Vol. 2009 (Spring), pp. 139-200.

Seater, J. J. (1993). 'Ricardian equivalence', *Journal of Economic Literature,* Vol. 31, pp. 142-190.

Sims, C. A. (1980). 'Macroeconomics and reality', *Econometrica,* Vol. 48, pp. 1Ű48.

FIGURE 1. Present values at time of passage for deficit-driven tax increases

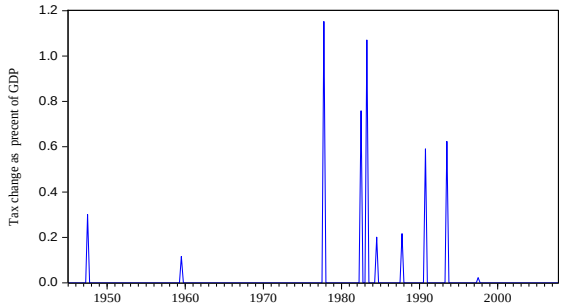

FIGURE 2. Estimated impact of the present value of an exogenous deficit-driven tax increase of 1% of GDP on GDP, as in equation (1) with no controls for lagged GDP growth and one and two standard-error confidence bands

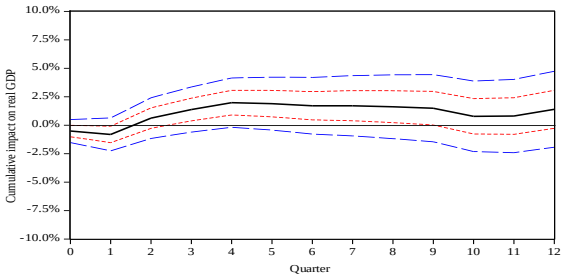

FIGURE 3. Estimated impact of the present value of an exogenous deficit-driven tax increase of 1% of GDP on GDP, as in equation (2) with controls for lagged GDP growth and one and two standard-error confidence bands (bootstrapped)

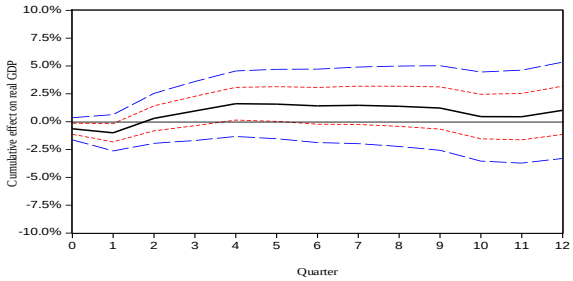

FIGURE 4. Post-1982:IV estimated impact of the present value of an exogenous deficit-driven tax increase of 1 % of GDP on GDP, as in equation (2) with controls for lagged GDP growth and one and two standard-error confidence bands (bootstrapped)

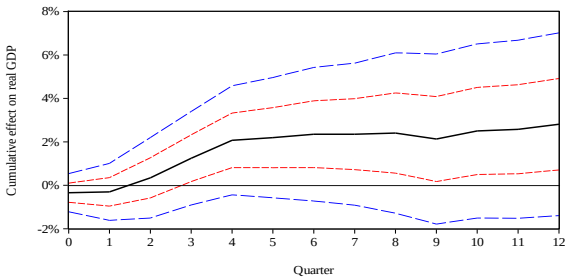

FIGURE 5a. Post-1982:IV estimated impact of an exogenous unanticipated deficit-driven tax increase of 1 % of GDP on GDP, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)

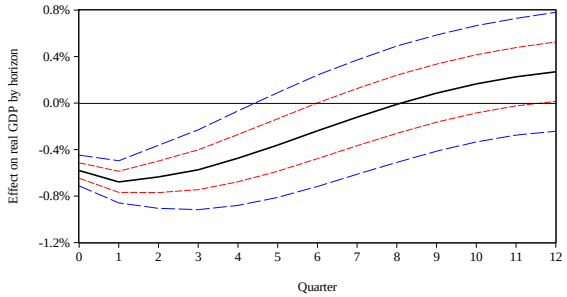

FIGURE 5b. Post-1982:IV estimated cumulative effect of an exogenous unanticipated deficit-driven tax increase of 1 % of GDP on GDP, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)

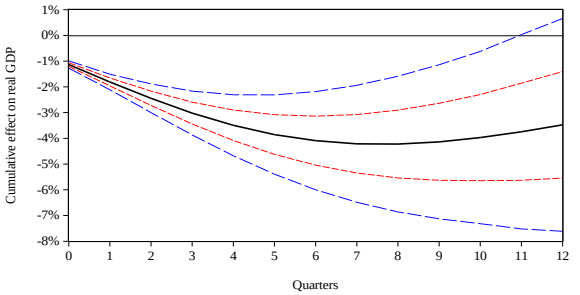

FIGURE 5c. Post-1982:IV estimated impact of an exogenous anticipated deficit-driven tax increase of 1 % of GDP on GDP, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)

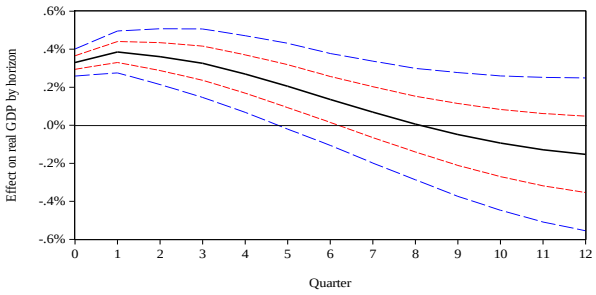

FIGURE 5d. Post-1982:IV estimated cumulative effect of an exogenous anticipated deficit-driven tax increase of 1 % of GDP on GDP, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)

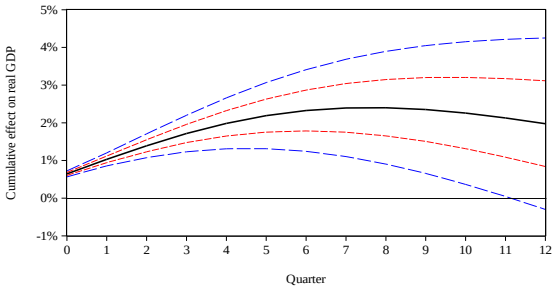

FIGURE 6a. Post-1982:IV estimated impact of an exogenous unanticipated deficit-driven tax increase of 1 % of GDP on consumption, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)

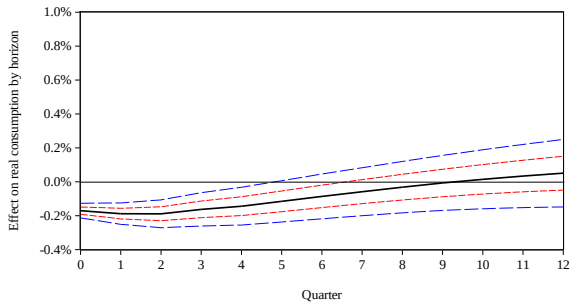

FIGURE 6b. Post-1982:IV estimated impact of an exogenous anticipated deficit-driven tax increase of 1 % of GDP on consumption, based on the five-variable VAR model with one and two standard-error confidence bands (bootstrapped)