Shafik Hebous - Tom Zimmermann

Revisiting the Narrative Approach of Estimating Tax Multipliers

Non-Technical Summary

A causal interpretation of tax policy changes on output is challenging for a number of reasons. First, a simple regression might yield biased estimates of the multiplier due to omitted variables. Second, causality can be reversed as spending and tax changes might be a response to changes in output. Third, even if these two concerns can be overcome, tax changes might have been anticipated, leading to output adjustments ahead of the tax change.

The literature mainly uses two different approaches to identify the causal effect of fiscal policy changes on output. One approach uses structural VAR models and imposes timing assumptions on the macroeconomic process arguing for example that taxes do not react to output within a quarter. The ability of this approach to identify a causal tax policy effect is heavily debated.

The second identification strategy was pioneered by Romer and Romer (2010) and attempts to directly identify the exogenous discretionary component in the tax series (a "narrative" measure). This approach has gained increasing popularity. A number of recent studies regress a narratively identified measure of a macroeconomic policy shock directly on an outcome variable.

In this paper, we argue that the narrative approach can be viewed as the reduced-form regression of an instrumental variable approach in which the narrative time series is used as an instrument for an endogenous series of interest. This motivates evaluating the validity of narrative measures through the lens of a randomized experiment. We apply our framework in a dynamic setting by extending the direct projection method of Jordà (2005) to allow for instrumental variables. In addition to being transparent, the method has the benefit that it is easy to incorporate weak instruments and to compute the appropriate confidence bands.

We use four recently constructed narrative measures of tax shocks by Romer and Romer (2010), Cloyne (2013), and Mertens and Ravn (2012). We document three empirical observations. First, for the US shock measures, we find several confounding factors indicating that a number of macroeconomic variables differ markedly between episodes of shock and episodes of no-shocks. Second, these narrative variables display low correlation with cyclically adjusted tax revenues. Third, the estimated effect of any of the considered narrative tax measures (with one exception) on output growth is insignificant when we use them as instruments.

The results of the dynamic analysis, and after correcting for weak instruments, show that using any of the considered narrative tax measures as an instrument for cyclically adjusted tax revenues yields tax multiplier estimates that are indistinguishable from zero. We conclude that the literature currently understates the uncertainty associated with quantifying the tax multiplier.

Revisiting the Narrative Approach of Estimating Tax Multipliers

Shafik Hebous[*]and Tom Zimmermann[†]

March, 2015

Abstract

A number of recent studies regress a "narratively" identified measure of a macroeconomic shock directly on an outcome variable. In this note, we argue that this approach can be viewed as the reduced-form regression of an instrumental variable approach in which the narrative time series is used as an instrument for an endogenous series of interest. This motivates evaluating the validity of narrative measures through the lens of a randomized experiment. We apply our framework to four recently constructed narrative measures of tax shocks by Romer and Romer (2010), Cloyne (2013), and Mertens and Ravn (2012). All of them turn out to be weak instruments for observable measures of taxes. After correcting for weak instruments, we find that using any of the considered narrative tax measures as an instrument for cyclically adjusted tax revenues yields tax multiplier estimates that are indistinguishable from zero. We conclude that the literature currently understates the uncertainty associated with quantifying the tax multiplier.

Keywords: Narrative Approach, Fiscal Stabilization, Tax Multiplier, Weak Instruments

JEL Classification: E62, H30, E69, C54

1 Introduction

The effect of a discretionary change in taxes or government spending on aggregate output is traditionally called "the multiplier". Estimating the multiplier is challenging because causality between changes in taxes and changes in output can run both ways and tax changes are often correlated with other economic developments. Romer and Romer (2010), hereinafter referred to as R&R, suggest an identification approach that relies on extracting the exogenous component of tax changes using published information in official policy documents ("narrative records"). R&R identify all significant legislated tax changes after World War II in the US, classify them as exogenous or endogenous changes, and estimate a peak cumulative tax multiplier of 3.

The narrative approach of R&R has gained increasing popularity, and it has been employed in several subsequent studies for different countries (e.g., the UK and Germany) and in different contexts (e.g., for measuring the effects of changes in government spending or social security benefits on economic output). The appendix provides a summary.

Studies that use the narrative approach typically specify a model in which the exogenous narrative variable is regressed directly on the outcome variable. The model can be a linear specification estimated by OLS (e.g., R&R, 2010) or a vector autoregression model with a narrative measure added as an exogenous variable (e.g. Cloyne, 2013).

In this note, we argue that, while narrative measures are commonly referred to as *shocks*, they should really be regarded as instruments for endogenous variables of interest.[1] This line of reasoning motivates questions that are typically investigated in instrumental variable (IV) applications: Is the instrument exogenous? Does it correlate strongly with the endogenous variable? Furthermore, we show that the reduced form specification typically employed in the literature only gives a correct estimate of the multiplier if the endogenous variable moves one-to-one with the narrative shock, a condition that is easily verifiable from the first stage regression of a two stage least squares estimation.

In order to apply our framework in a dynamic setting, we extend the direct projection method of Jordà (2005) to allow for instrumental variables. In addition to being transparent, the method has the benefit that it is easy to incorporate *weak* instruments and to compute the appropriate confidence bands.

As a specific application, we closely reexamine four narrative tax variables: All legislated tax changes in the US by R&R, two variables by Mertens and Ravn (2012) that are derived

---

[1]This point was first developed by Stock and Watson (2012), who state that narrative measures "are not the full shock series, rather they measure (typically with error) an exogenous component of the shock, so that the constructed series is correlated with the shock of interest but not with other shocks."

from R&R and that distinguish between anticipated and unanticipated tax changes, and the measure of all legislated tax changes in the UK by Cloyne (2013). These variables provide a homogenous context and were all regressed directly on output changes in the corresponding studies.

We document three empirical observations. First, for the US shock measures, periods in which shocks occur differ markedly from those in which no shocks occur. Shock periods typically have higher long- and short-term interest rates, higher unemployment and lower debt-to-GDP ratios. We also find that narrative tax shocks can be predicted by lagged macroeconomic variables in both the US and in the UK. Our final specifications therefore include many macroeconomic control variables, such that the instrument (the narrative shock) can be viewed as being "as-if" randomly assigned, conditional on observables.

Second, the shocks display low correlation with cyclically adjusted tax revenues, that is, they do not move one-to-one with the endogenous variable that we consider. As F-statistics of first-stage regressions are small in most specifications, we correct the confidence bands of the estimated coefficients for weak instruments as suggested by Staiger and Stock (1997). Unsurprisingly, we find stronger correlations in the first stage when we consider *all* legislated tax changes (the sum of changes that are categorized by the original authors as endogenous or exogenous) as the endogenous variable. Still, correlations remain small for the narrative measures of Mertens and Ravn (2012).

Third, the estimated effect of any of the considered narrative tax measures (with one exception) on output growth is insignificant when we use them as instruments. When instruments are weak, confidence bands can become very wide, always include 0 and frequently include multiplier values of -10 or lower. We conclude that the literature currently understates the uncertainty associated with estimating the tax multiplier.

Our contribution stresses the issue of the relevance of the narrative measure for endogenous variables of interest. As such, it is related to the external instruments approach of Stock and Watson (2012) and the discussion of weak external instruments in a VAR setup in Montiel Olea et al. (2012). Another distinct point that complements our discussion focuses on the exogeneity of a narrative variable. This route was taken by Jordà and Taylor (2013) who argue that narratively identified episodes of fiscal consolidations, as identified by IMF (2010) are not random events.[2] The issues of relevance and exogeneity, together, give a general guide of how to view the narrative approach to the estimation of macroeconomic effects. We discuss some broader implications for empirical macroeconomics in the conclusion.

---

[2]The exogeneity aspect of narrative variables was also debated in the context of monetary policy shocks (Leeper, 1997; Romer and Romer, 1997).

2 The Narrative Approach and an IV Perspective

2.1 Conceptual Framework

We start from the conceptual framework introduced in R&R that motivates their construction of a narrative tax shock. Abstracting from any dynamics for illustrative purposes, they suggest to estimate the tax multiplier from a simple regression of changes in output (Δy_t) on changes in taxes (ΔT_t).[3]

$$\Delta y_t = \alpha + \beta \Delta T_t + \epsilon_t. \tag{1}$$

In equation (1), β can be interpreted as the contemporaneous tax multiplier.[4] A causal interpretation of β, however, is challenging for a number of reasons. First, there could be omitted variables (e.g. policy rate changes) that would bias the coefficient. Second, tax changes might be a response to changes in output such that causality would be reversed. Third, even if these two concerns can be overcome, tax changes might have been anticipated, leading to output adjustments ahead of the tax change. In all of these cases, one would expect ΔT_t to be correlated with the error term ϵ_t in equation (1).

R&R argue that changes in taxes can be decomposed into two components, an exogenous one and an endogenous one. The first component captures exogenous discretionary fiscal policy actions. The second component includes non-fiscal policy effects and automatic changes in the fiscal series over the business cycle. Additionally, it includes fiscal policy changes that are taken in reaction to the business cycle.

More specifically, let changes in taxes be

$$\Delta T_t = \sum_{i=1}^{K} b_t^i \epsilon_t^i + \sum_{j=1}^{L} \omega_t^j, \tag{2}$$

with ϵ_t being the sum of all endogenous disturbances in a quarter t, i.e., $\epsilon_t = \sum_{i=1}^{K} \epsilon_t^i$.

The first term on the right-hand side of equation (2) depicts endogenous changes of ΔT_t and the second term depicts exogenous tax changes. Denoting exogenous tax changes as $\omega_t \equiv \sum_{j=1}^{L} \omega_t^j$, plugging equation (2) in equation (1), and re-arranging terms yields

$$\Delta y_t = \alpha + \beta \omega_t + u_t, \tag{3}$$

---

[3]When they proceed to estimate the model, they include lagged values of output changes and other variables as well.

[4]A dynamic analysis often includes lagged values of the fiscal variable. Then, the cumulative multiplier is the sum of the coefficient associated with contemporaneous ΔT_t and the coefficients on the included lagged values of ΔT_t.

4

where $u_t = \sum_{i=1}^{K}(1 + \beta b_t^i)\epsilon_t^i$.

If ω_t is truly exogenous (uncorrelated with u_t), β is identified from equation (3) with a causal interpretation as tax multiplier. This framework motivates R&R's construction of a measure of exogenous tax changes (the *narrative* measure) from official sources such as the *Economic Report of the President* that helps to categorize the motivation behind each single reported tax change. When they estimate equation (3) and variants of it that include dynamics and controls, they find that the tax multiplier peaks at a value of around 3 after 2-3 years with one standard error confidence bands between 2 and 4.

2.2 Narrative Measures as Instruments

The narrative measure of R&R is commonly referred to as a *shock*, but, more precisely, can be regarded as an instrument for an endogenous variable of interest. As noted by Stock and Watson (2012), narrative measures "are not the full shock series, rather they measure (typically with error) an exogenous component of the shock, so that the constructed series is correlated with the shock of interest but not with other shocks."

This interpretation motivates the estimation of equation (1) and using ω_t as an instrument for the endogenous variable ΔT_t in the first-stage regression

$$\Delta T_t = \delta_0 + \delta_1 \omega_t + \upsilon_t. \tag{4}$$

The success of this two-step approach will typically depend on the strength of the correlation between the instrument and the endogenous variable. This strength can be assessed by comparing the F-statistic of the first stage regression to the suggestions in Stock and Yogo (2005) or Montiel Olea and Pflueger (2013).

Furthermore, using the simple decomposition in equation (2), the magnitude of δ_1 in the first stage regression (4) can be interpreted meaningfully: Note that $\delta_1 = \frac{\text{Cov}(\Delta T_t, \omega_t)}{\text{Var}(\omega_t)}$ and, as,

$$\begin{aligned} Cov(\Delta T_t, \omega_t) &= Cov(\epsilon_t + \omega_t, \omega_t) \\ &= Cov(\epsilon_t, \omega_t) + Var(\omega_t), \end{aligned} \tag{5}$$

it follows that

$$\delta_1 = \frac{Cov(\epsilon_t, \omega_t)}{Var(\omega_t)} + 1. \tag{6}$$

The first component on the RHS of equation (6) measures the correlation of the shock of interest, ω_t, with other shocks. Under the assumption that the shock of interest is accurately identified, that is, it is uncorrelated with other shocks, δ_1 should therefore be equal to 1, which we summarize in the following corollary:

Corollary 1 *If exogenous tax changes are accurately identified and if equation (2) holds,*

$$\frac{Cov(\Delta T_t, \omega_t)}{Var(\omega_t)} = 1. \tag{7}$$

This corollary is nothing more than an identity if the endogenous measure of total taxes is constructed as the sum of endogenous and exogenous tax changes. In that case, total taxes will move one-to-one with each component of tax changes. Often, though, the components are unknown and the object of interest is an easily measurable endogenous series of taxes as, e.g. a series of cyclically adjusted tax revenues. If δ_1 is low in that case, it would raise the question whether the proposed shock is well reflected in the endogenous variable.

Another way to connect our IV framework to the approach of R&R is to note that their specification (in equation (3)) is the so-called *reduced-form* of an IV estimator. The reduced-form regression is a regression of the outcome variable of interest (here: output growth) on the instrument. The coefficient in the reduced-form regression needs to be divided by the coefficient in the first stage regression to account for instrument strength, and only if $\delta_1 = 1$, as in our corollary, is the reduced form estimate equal to the structural estimate β in equation (1).[5]

---

[5]More specifically, let the reduced form regression be

$$\Delta y_t = \gamma_0 + \gamma_1 \omega_t + \eta_t, \tag{8}$$

where η_t is an error term. It can be shown that

$$\beta = \frac{\gamma_1}{\delta_1}, \tag{9}$$

and β, δ_1 are defined as in equations (1) and (4). It follows that $\beta = \gamma_1$ if and only if $\delta_1 = 1$.

2.3 The Implementation of IV Estimation in a Dynamic Setting

Our empirical implementation of the IV framework accounts for dynamics using the local projection approach introduced in Jordà (2005). He suggests to estimate impulse response functions (IRFs) via regressions of the form

$$Y_{t+h} = a^h + b^h \omega_t + \sum_{j=0}^{K} \phi_j^h \Delta Y_{t-j} + u_t^h, \forall h \in \{0, \ldots, H\}. \tag{10}$$

The reduced form IRF is then directly given by the set of coefficients $\{b^h\}_{h=0}^{H}$. Jordà (2005) shows that the resulting IRF from (10) is identical to that obtained from autoregressive models. One advantage of the local direct projections representation is its straightforward extension to an IV model by considering the form

$$Y_{t+h} = \alpha^h + \beta^h \Delta T_t + \sum_{j=0}^{K} \gamma_j^h \Delta Y_{t-j} + e_t^h, \forall h \in \{0, \ldots, H\}, \tag{11}$$

and the exogenous narrative tax measure ω_t is used as an instrument for ΔT_t as in equation (4). The IV IRF is then directly given by the set of coefficients $\{\beta^h\}_{h=0}^{H}$.

As we frequently encounter cases in which the narrative shock is only weakly correlated with the endogenous variable in our application below, we correct the confidence bands for weak instruments in a standard fashion. In particular, we compute confidence bands by inverting the Anderson-Rubin test statistic as motivated in Staiger and Stock (1997).

3 Empirics

3.1 Empirical Regularities

We consider four recently constructed narrative tax variables: The seminal measure of R&R capturing all exogenous legislated tax changes in the US, two measures by Mertens and Ravn (2012) that distinguish between anticipated and unanticipated tax changes in the US and that are derived from R&R, and the measure of all exogenous legislated tax changes in the UK by Cloyne (2013). All data are described in detail in the online appendix.

3.1.1 Instrument Validity: Potential Confounders

IV estimation requires that the instrument is "as-if" randomly assigned across observations. In our application, observations are different time periods and we investigate whether the

macroeconomic environment is similar across periods in which shocks occur (we label those periods "treated") and in those in which no shocks occur (we label those periods "control").

In the appendix, we apply two standard tests to assess the exogeneity of the narrative shocks. First, we check for possible imbalances of the covariate distributions in the treatment and the control group. Results show that the macroeconomic variables are unevenly distributed across the two groups for the US narrative tax measures. In particular, we find that long- and short-term interest rates and the unemployment rate are typically higher and the debt-to-GDP ratio is typically lower in treated periods. The covariate distributions look balanced for Cloyne (2013)'s UK narrative measure. Second, we check whether the occurrence of narrative shocks can be predicted by past values of macroeconomic variables. Here, we find that narrative shocks in both the US and in the UK are predictable by past interest rates.

To shield against variables that would confound the causal interpretation of the tax multiplier, in the estimations below, we include a set of macroeconomic variables and their lags . The assumption is, thus, that the narrative shocks are "'as-if'" randomly assigned, conditional on the variables that we include. We note that unobserved variables could still confound the interpretation.

3.1.2 Instrument Relevance

Table 1 reports estimates of the first-stage regression in equation (4) for all four narrative tax measures. For each of the considered narrative exogenous variable, we examine two different endogenous series: The series of all legislated tax changes that are identified based on the narrative records and the ratio of real cyclically adjusted revenues to GDP.[6] First, note that the estimated coefficient δ_1 is significant throughout almost all specifications, regardless of the endogenous variable. At the same time, however, if we consider the endogenous series to be the ratio of real cyclically adjusted revenues, δ_1 is below 1 for all narrative shocks. For example, in the case of the unanticipated tax changes, δ_1 is equal to 0.2, and even for better performing measures, δ_1 does not exceed .65. These results are not in line with the suggested decomposition given by (2). However, when considering all legislated tax changes as the underlying endogenous series, the ratio for the variable of Cloyne (2013) is exactly 1 and for R&R very close to 1. The variables of Mertens and Ravn (2012) give small ratios. Additionally, Table 1 reports the first-stage regression results of a dynamic model including lagged output, lagged values of the exogenous narrative measure and covariates such as the

---

[6]In the case of anticipated and unanticipated tax changes, the series of all legislated tax changes is that of R&R since the episodes are a subset of the original R&R narrative variable.

policy rate. The estimated coefficients on the excluded instrument are rather similar to those obtained from the model without accounting for confounders.

Deviations of the covariance-variance ratio from 1 can be interpreted in different ways. Notably, the shocks ω_t might not be precisely measured or the proposed decomposition of total taxes given in (2) might not hold; for instance, changes in ω_t might not translate one-to-one into total tax changes ΔT_t. The most serious concern would be that the narrative series are not related to the endogenous series of interest. In any case, the result suggests that estimating the tax multiplier from the reduced form equation (3) alone might not be warranted, as the narrative shocks do not map one-to-one into tax changes.

To assess the strength of the first stage regression, table 1 reports the $F-$statistic of the model. As a rule of thumb, if the $F-$statistics of the first-stage regression is below 10 then the set of instruments is considered to be weak (Staiger and Stock, 1997). Table 1 shows that the $F-$statistics are mostly smaller than 10 (except for R&R's narrative measure) when we use cyclically adjusted revenues in the first-stage. The $F-$statistics are greater than 10 for all variables when we use the narrative endogenous series in the first-stage (except for the case of the anticipated tax measure). We take this into account by using weak IV robust confidence bands as described in the previous section.

3.2   IV Estimation Results

Figure (1) plots the IRFs of output following a tax shock of 1% of GDP identified based on narrative measures and using our IV method. Since results in Table 1 suggest that narrative measures tend to be weakly correlated with the endogenous regressor, we compute weak-IV robust 95% confidence intervals, depicted as the gray regions. For comparison, we plot the IRFs obtained from the reduced form (10) with the corresponding 95% non-corrected confidence bands shown as dotted lines. All estimations include 12 lags. All regressions for the US control for the monetary policy rate, AAA corporate bond rate, corporate spread, defense spending, and the unemployment rate.

The findings are summarized as follows. First, when the instrument is strong, as in panels (a) and (c), results from the reduced-form regression and from the IV method are very similar. Using the narrative measures of R&R and Cloyne as instruments for all legislated tax changes yields similar patterns to the original results reported in their papers. For R&R, the IRF is significant at the 95% confidence level for intermediate horizons of three to six quarters. Using the UK data of Cloyne (2013), results are statistically weaker and again do not depend on the methodology. Second, when the instrument is weak, as in all other panels, confidence bands for the tax multiplier become very wide, and typically include 0 and often values of

-10 or larger in magnitude. This is the case for all estimates that use cyclically adjusted tax revenues as the endogenous variable, regardless of which narrative measure is used as instrument. It is also the case when we use all legislated tax changes as the endogenous variable and consider the unanticipated and anticipated shocks of Mertens and Ravn (2012) as instruments.

One caveat regarding the question: What is the endogenous series? The answer generally depends on the research question, and there might be different views on this. For instance, Mertens and Ravn (2013) use the narrative measure as an instrument for the series of reduced form innovation shocks obtained from a vector autoregression model. The message of this note is independent of this discussion and it stresses two issues. Firstly, for the interpretation of the results, the research design has to be explicit about the endogenous series. Secondly, if it is weakly correlated with the exogenous narrative measure, one must use corrected confidence bands for the weakness of the instrument.

4   Concluding Remarks

In this note, we interpret narrative shocks as instruments for endogenous variables of interest. We apply the framework to the estimation of tax multipliers and we find that estimates are often insignificant, due to two reasons: First, the statistical relation between output changes and narrative shocks is relatively weak to begin with (recall that the literature mostly reports 68% confidence bands). Second, narrative tax shocks often show low correlations with the endogenous tax measure. This is the case for all shocks when the endogenous variable is the series of cyclically adjusted tax revenues, and it is true for the shock series of Mertens and Ravn (2012) when the endogenous variable is the series of all legislated tax changes. Accounting for the weak correlation in the first stage generally leads to wide confidence bands for estimates of the tax multiplier. Our results indicate that the literature currently understates the uncertainty associated with estimating the tax multiplier.

While this note discusses the estimation of the tax multiplier using narrative shocks, the framework is more general and invites empirical macroeconomists to think about applications from an experimental design perspective. What is the underlying experiment that allows a causal interpretation of the estimate? What is the underlying endogenous variable? Or, somewhat more unusual, what is the minimum effect that one could detect in a sample of T observations? We hope that this view of thinking about empirical macroeconomics will prove useful in future applications.

References

Cloyne, J. (2013). Discretionary tax changes and the macroeconomy: new narrative evidence from the united kingdom. *American Economic Review 103*(4), 1507–1528.

IMF (2010, September). *Recovery, Risk, and Rebalancing*, Chapter 3. World Economic Outlook.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Jordà, Ò. and A. M. Taylor (2013). The time for austerity: estimating the average treatment effect of fiscal policy. Technical Report 19414, National Bureau of Economic Research Working Paper.

Leeper, E. M. (1997). Narrative and var approaches to monetary policy: Common identification problems. *Journal of Monetary Economics 40*(3), 641–657.

Mertens, K. and M. O. Ravn (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy 4*(2), 145–81.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review 103*(4), 1212–1247.

Montiel Olea, J. L. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2012). Inference in structural vars with external instruments. Technical report, mimeo.

Romer, C. D. and D. H. Romer (1997). Identification and the narrative approach: A reply to leeper. *Journal of Monetary Economics 40*(3), 659–665.

Romer, C. D. and D. H. Romer (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review 100*, 763–801.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007–09 recession. *Brookings Papers on Economic Activity 2012*(1), 81–135.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*.

Figure 1: IRFs of output growth to changes in different tax measures

(a) R&R Exo. as IV for RR All

(b) R&R Exo. as IV for Cyc. Adju.

(c) Cloyne Exo. as IV for. Cloyne All

(d) Cloyne Exo. as IV for Cyc. Adj.

(e) M&R Unanti. as IV for RR All

(f) M&R Unanti. as IV for Cyc. Adj.

(g) M&R Anti. as IV for RR All

(h) M&R Anti. as IV for Cyc. Adj.
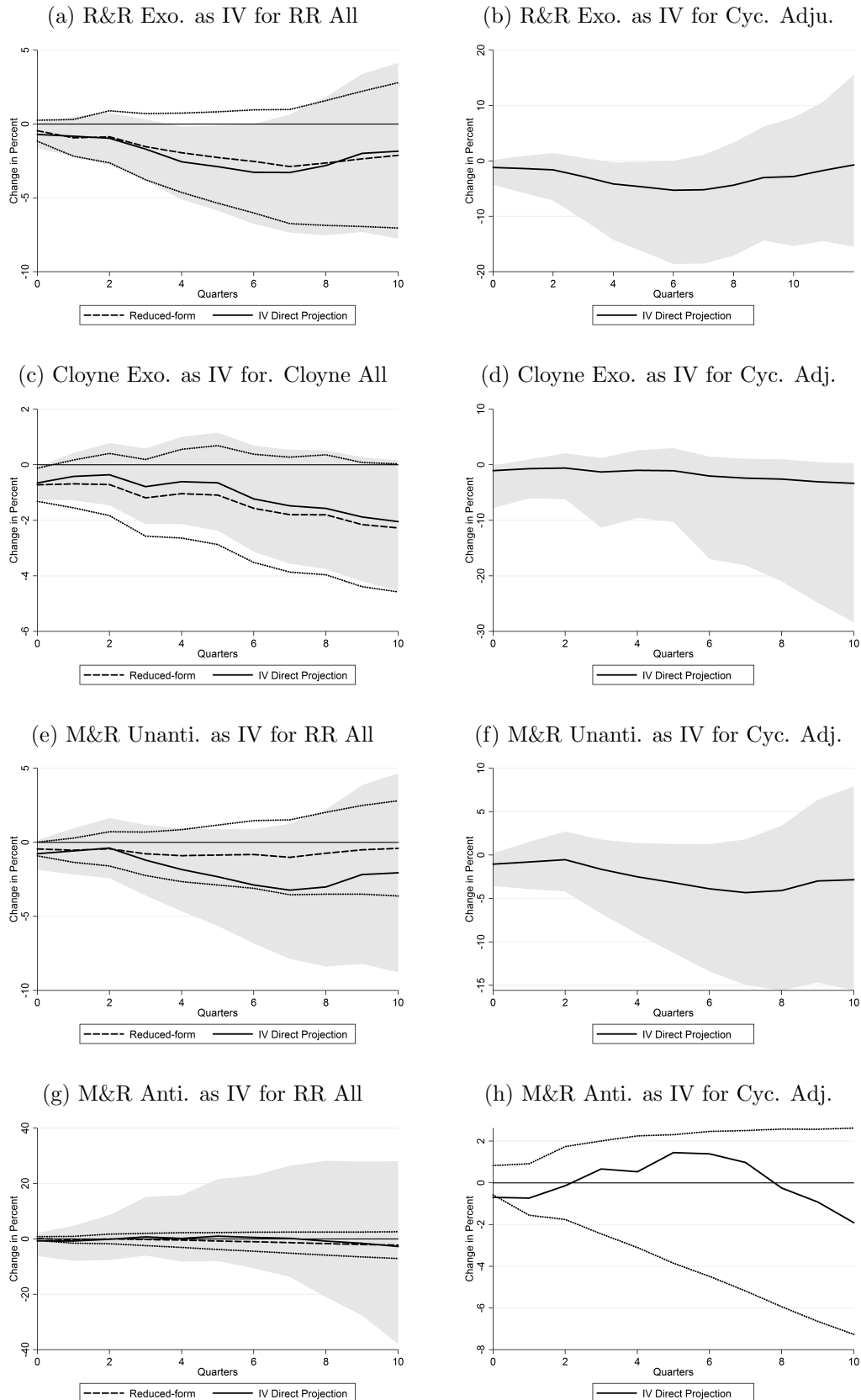
Note: IV IRFs are based on equation (11) whereas reduced form IRFs are based on equation (10). Grey regions are 95% confidence bands corrected for the weak instrument problem. Dotted curves are reduced form 95% confidence bands. "RR Exo." denotes the exogenous variable of R&R (2010). "Cyc. Adj." denotes the ratio of real cyclically adjusted revenues. "Cloyne Exo." denotes the exogenous variable of Cloyne (2013). "M&R Unanti." and "M&R Anti." stand for unanticipated and anticipated legislated tax changes of Mertens and Ravn (2012), respectively.

Table 1: Statistics from First-Stage Regressions

Endogenous Series	Specification	Sample Period	ΔT_t is Cyclically Adj. Taxes				ΔT_t is: All Legislated Changes			
			δ_1	$t-stat.$ $(prob>t)$	R^2	$F-stat.$	δ_1	$t-stat.$ $(prob>t)$	R^2	$F-stat.$
R&R (2010)	No controls	1947q2-2007q4	0.502	3.45 (0.00)	0.07	11.90	0.958	26.93 (0.00)	0.360	725.34
	Controls	1957q3-2004q3	0.542	2.85 (0.00)	0.45	8.48	0.861	7.72 (0.00)	0.70	71.42
Cloyne (2013)	No controls	1955q2-2009q4	0.432	1.40 (0.16)	0.01	1.96	1	5.4 (0.00)	0.24	29.10
	Controls	1958q2-2006q4	0.631	2.24 (0.027)	0.18	5.16	1.02	7.37 (0.00)	0.29	57.05
MR (2012), Unanticipated	No controls	1947q2-2006q4	0.198	2.21 (0.02)	0.03	4.88	0.417	8.52 (0.00)	0.20	72.50
	Controls	1957q3-2004q3	0.339	2.96 (0.00)	0.46	9.59	0.473	6.7 (0.00)	0.67	47.53
MR (2012), Anticipated	No controls	1947q2-2006q4	0.332	2.31 (0.02)	0.02	5.35	0.451	2.67 (0.02)	0.06	7.11
	Controls	1957q3-2003q3	0.371	1.88 (0.06)	0.44	3.10	0.292	2.15 (0.03)	0.55	5.91

Note: Results are based on regressions of the form

$$\Delta T_t = \delta_0 + \delta_1 \omega_t + \upsilon_t,$$

with and without control variables. The variable ω_t is the narrative variable that is labeled as *exogenous* in the corresponding study given in the first column of this table. The variable ΔT_t is the endogenous series. We consider two potential endogenous series. 1) *Cyclically Adj. Taxes*: The real cyclically adjusted tax revenues as a percent of real GDP. 2) *All legislated tax changes* as identified by the narrative approach (i.e., endogenous legislated changes plus exogenous legislated changes). Specifications without controls do not include any lagged values of the narrative variable any covariates whereas specifications with controls are dynamic models including 12 lags as described in (11) and (4), and in the case of the US include the following controls: The monetary policy rate, AAA corporate bond rate, corporate spread, defense spending, and the unemployment rate. The sample period differs for the dynamic models due to the inclusion of 12 lags and the availability of the policy rate for the US (starting from 1954q2).

Online Appendix for the Paper:
Revisiting the Narrative Approach of Estimating Tax Multipliers

Shafik Hebous[*]and Tom Zimmermann[†]

March, 2015

---

[*]Goethe University Frankfurt, hebous@wiwi.uni-frankfurt.de
[†]Harvard University, tzimmerm@fas.harvard.edu.

A   Online Appendix

A.1   Summary of Recent Studies that Use Narrative Measures

Table A.1: Summary of Recent Studies that Use Narrative Measures

Studies that construct a narrative variable

Study	Exogenous Narrative Variable	Country	Sample Period
Studies on tax multipliers			
Romer and Romer (2010)	all legislated taxes	US	1945-2006
Cloyne (2013)	all legislated taxes	UK	1948-2009
Mertens and Ravn (2012)	unanticipated taxes	US	1947-2006
Mertens and Ravn (2012)	anticipated taxes	US	1947-2006
Mertens and Ravn (2013)	personal taxes	US	1950-2006
Mertens and Ravn (2013)	corporate taxes	US	1950-2006
Hayo and Uhl (2014)	all legislated taxes	Germany	1974-2010
Guajardo et al. (2014)	all fiscal consolidation actions	panel of 17 OECD countries	1978-2009
Other studies			
Bahaj (2014)	sovereign risk premia	Eurozone countries	2009-2013
Dominguez and Shapiro (2013)	policy and financial shocks	US and EU	2008-2012
Ramey (2011)	defense spending news	US	1939-2008
Romer and Romer (2014)	social security transfers	US	1952-1991

Studies that use an existing narratively identified variable in their applications

Study	Used Variable	Application
Feyrer and Shambaugh (2012)	Romer and Romer (2010)	effects of tax policy on the current account
Perotti (2012)	Romer and Romer (2010)	compare automatic stabilizers and discretionary policy effects
Bluedorn and Leigh (2011)	Guajardo et al. (2014)	effects of fiscal consolidation on the current account
Alesina et al. (2014)	Guajardo et al. (2014)	effects of multiple-period fiscal consolidation on output

Note: This table lists a number of recent studies that use the narrative approach to identify macroeconomic effects, and is not meant to be an exclusive survey of the literature. The upper group of studies constructs variables from the narrative records to estimate effects on output and other macroeconomic aggregates. The lower-panel group of studies readily uses existing narrative variables in their applications.

## A.2 Summary of Considered Narrative Measures

We consider four narrative tax variables: The seminal measure of R&R capturing all exogenous legislated tax changes in the US, two subset variables by Mertens and Ravn (2012) that distinguish between anticipated and unanticipated tax changes in the US by labeling a tax liability change as unanticipated if the implementation lag does not exceed 90 days, and the measure of all exogenous legislated tax changes in the UK by Cloyne (2013). Figure A.1 plots the considered narrative variables. Since legislated tax changes are by construction discrete events, these tax series include a number of zero-observations indicating no tax actions in these quarters. Table A.2 in the appendix reports means and standard deviations of these variables.

In the specific context of tax multipliers the ultimate endogenous series of interest can be debated. Alesina and Ardagna (2010) use cyclically adjusted tax series, which arguably offers a clear observable tax measure. Others argue against it without providing an alternative endogenous series; e.g., IMF (2010). Several studies that use the narrative identification approach refrain from clearly defining the underlying endogenous series. R&R and Cloyne (2013) are exceptional with this regard, and therefore we present our idea using their variables. Thus, we use two different variables as the underlying endogenous series of tax changes: The sum of all legislated tax changes and the series of real cyclically-adjusted tax revenues. Our aim is to provide a general message, not specific to the context of taxes, stating that the endogenous series is an essential part of a meaningful narrative identification approach.

Further, figure A.1 presents the narrative measures of Mertens and Ravn (2012). These measures addresses the challenge that there often exists a lag between the announcement of fiscal policy changes and the date of implementation. The implementation lag enables agents in the economy to anticipate the fiscal action and react immediately at the announcement date and before the implementation. If the announcement is in t and the implementation in $t+1$ or $t+2$, this can trigger a change in Δy at time t even though the change in T has not yet occurred and will actually occur in the data in a later period. Therefore, anticipation of fiscal actions makes an OLS estimate of β biased because the fiscal variable will be correlated with future values of the error term. Mertens and Ravn (2012) distinguish between anticipated and unanticipated legislated tax changes in the US (the last two panels of figure, A.1) by labeling a tax liability change as unanticipated if the implementation lag does not exceed 90 days.

In addition to the narrative series, we use a number of additional macroeconomic time series as control variables for the US.

Figure A.1: Narrative Tax Changes
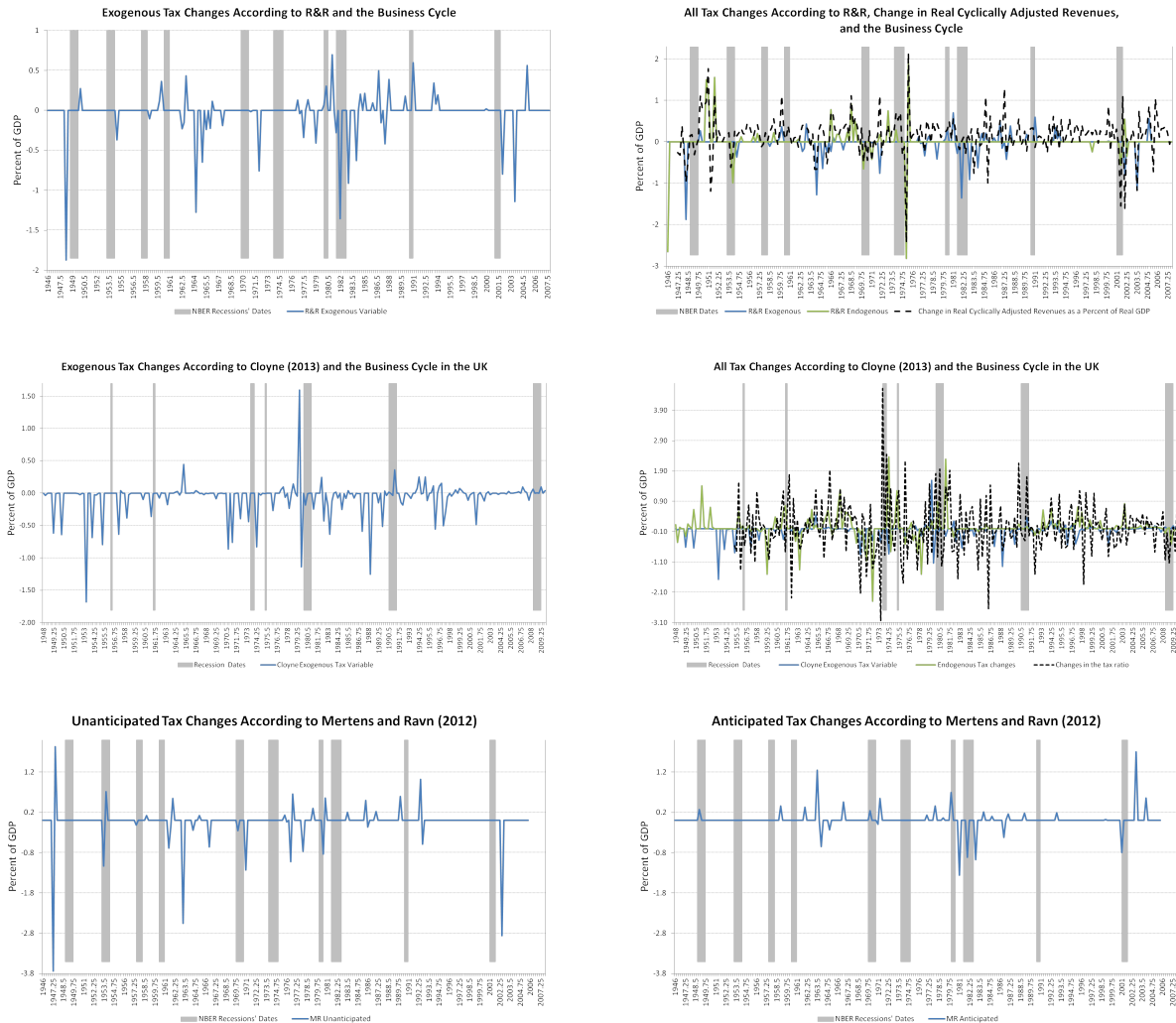
Note: Exogenous tax changes are legislated tax changes that are identified by the narrative records and classified as exogenous by the original authors (i.e. Romer and Romer (2010) for the US and Cloyne (2013) for the UK). Endogenous tax changes are legislated tax changes that are identified by the narrative records and classified as endogenous by the original authors. Unanticipated and anticipated legislated tax changes are taken from Mertens and Ravn (2012). The gray bars indicate episodes of recessions. US dates of recessions are taken from the NBER whereas the dates for the UK are based on authors' own calculation and defined as two consecutive quarters of negative economic growth.

Table A.2: Summary of Considered Narrative Measures

Study	Measure	No. of Non-zero Observations	Mean (SD)
Romer and Romer (2010)	all legislated taxes	45	-0.03 (0.24)
Romer and Romer (2010)	all exogenous legislated taxes	69	-0.02 (0.41)
Cloyne (2013)	all legislated taxes	124	-0.06 (0.25)
Cloyne (2013)	all exogenous legislated taxes	171	-0.02 (0.47)
Mertens and Ravn (2012)	unanticipated taxes	31	-0.04 (0.42)
Mertens and Ravn (2012)	anticipated taxes	27	0.01 (0.22)

Note: All listed tax variables are constructed based on the narrative records. The studies of Romer and Romer (2010) and Mertens and Ravn (2012) are for the US whereas Cloyne (2013) is for the UK. The variables are expressed as a percent of GDP. In the econometric analysis, the sample in Romer and Romer (2010) spans from 1947q1 to 2006q4 whereas the sample in Cylone (2013) spans from 1955q1 to 2009q4.

A.3   Additional Variables

The sources of the narrative variables are: Romer and Romer (2010) for exogenous and endogenous legislated taxes changes in the US. Cloyne (2013) for exogenous and endogenous legislated taxes changes in the UK. Mertens and Ravn (2012) for unanticipated and anticipated legislated taxes changes in the US.

Concerning the macroeconomic data for the US, the change in real GDP is taken from Romer and Romer (2010). The variable house prices is taken from FRED: ASPUS. The CPI is taken from FRED: CPIAUCSL. Government spending is real federal spending expressed as the change from the last quarter and it is taken from Ramey (2011). Defense spending is real defense spending expressed as the change from the last quarter and it is taken from Ramey (2011). News spending is the nominal present value of the news variable of Ramey (2011). The variable Debt/GDP is the ratio of public debt to GDO obtained from FRED: GFDEGDQ188S. The policy rate is obtained from FRED: FEDFUNDS. The short-term government bond rate is from FRED: TB3MS whereas the long-term government bond rate is from FRED: GS10. AAA corporate bond rate is from FRED: AAA whereas BAA corporate bond rate is from FRED: BAA. SP500 is the stock price index taken from FRED: SP500. SPXD is a stock price index obtained from Global Financial Data (ticker: SPXD). Consumer confidence index is taken form FRED: UMSCENT & UMSCENT1.

For the UK, real GDP is taken from ONS as in Cloyne (2013). The policy rate is taken from Cloyne (2013). The corporate bond yield is taken from Global Financial Data (ticker: INGBRW). The financial index FTASD is taken from Global Financial Data (ticker: FTASD). The long-term government bond rate is obtained from Global Financial Data: 20 year government bond yield (ticker: IGGBR20D). The short-term government bond rate is obtained from Global Financial Data: 3-month treasury bill yield (ticker: ITGBR3D). Government spending is the variable: Government final consumption expenditure, volume; obtained form the OECD Economic Outlook vol. 93. Regarding the global variables, the first oil price index is Brent crude oil taken from Global Financial Data (ticker: BRT_D). The second oil price index is WTI index obtained from Global Financial Data (ticker: WTC_D).

A.4   Potential Confounding Factors

One concern for the estimation is that there might be differences, in terms of other macroeconomic variables, between episodes of exogenous tax policy actions (treatment group) and episodes without these actions (control group), which would imply that the shocks are not "as-if" randomly assigned across periods. The experimental design literature typically reports two standard tests which we apply below.

We compute average values and standard deviations of a number of potential covariates for the treatment and control groups as identified by R&R, and report the differences between the means. Table A.3 displays the results. We note that in such an exercise, the t-statistic is not an informative

indicator since it is partially driven by the sample size. Instead, we document normalized mean differences in the last column. While a larger t-statistic could reflect larger sample size, a larger normalized difference unambiguously indicates a more unbalanced covariate distribution between the two groups. As a rule of thumb, a difference between average covariates values of 0.25 standard deviations or more suggests a violation of the exclusion restriction.

Table A.3 reveals that financial conditions tend to be different depending on the type of the episode. The normalized differences are larger than 0.25 standard deviations in the case of the SP500 index, AAA and BAA corporate bond rates, and long-term and short-term government bond yields. The monetary policy rate appears to be similar. Among the considered macroeconomic variables, the difference in normalized means of changes of the unemployment rate is visibly large (0.4 standard deviations). Also, some fiscal variables such as the federal debt ratio and changes in defense spending tend to be different depending on the treatment status. The spending news variable of Ramey (2011) seems balanced between both types of periods, though.

Looking at Cloyne (2013)'s measure for the UK, table A.3 shows that all considered covariates do remarkably well in this regard. Normalized differences in covariates' average values are rather small.

Additionally, table A.3 shows that differences between episodes of anticipated tax changes and no anticipated tax actions as identified by the narrative measures of Mertens and Ravn (2012) are pervasive. For example, normalized differences are above 0.5 standard deviations in the case of long-term government bond rate and AAA and BAA corporate bond rates. Overall, with the exception of Cloyne (2013), the results of this exercise suggest that episodes of tax changes are generally different in some aspects from episodes of no tax changes. This means that unless the research design takes differences in the economic environment into account, there is no guarantee that the obtained results apply in other macroeconomic configurations (e.g., low unemployment, low bond spread). We therefore include variables that display large imbalances as control variables in our empirical framework.

As a further test, we check whether episodes of non-zero tax shocks are predictable by *lagged* macroeconomic variables. We define the dummy variable S_t that equals 1 if the tax shock is non-zero in period t, we estimate the binary regression model

$$P(S_t = 1) = \Phi(d_1 X_{t-1} + d_2 X_{t-2}), \tag{1}$$

for different macroeconomic variables X_t, and we test the hypothesis that both d_1 and d_2 equal 0.

Table A.4 provides results. We reject the null that the narrative measures for both the US and the UK are not predictable by macroeconomic variables. Instead, we find evidence that past values of interest rates or interest rate spreads and the unemployment rate forecast the occurrence of narrative shocks. We therefore include lagged macroeconomic variables in our estimations as additional controls.

Table A.3: Confounding Factors

| | Romer and Romer (2010) | | | | | Cloyne (2013) | | | | |
| | Tax Changes | | No Tax Changes | | diff/SD | Tax Changes | | No Tax Changes | | diff/SD |
Variable	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Financial Variables										
SP500 / FTASD for Cloyne (2013)	1.02	0.47	0.88	0.38	**0.34**	0.02	0.08	0.03	0.11	-0.11
SPXD	0.03	0.07	0.02	0.08	0.13					
AAA corporate bond rate	7.70	3.33	6.41	2.81	**0.44**	8.87	3.32	8.56	3.56	0.09
BAA corporate bond rate	8.71	3.73	7.29	3.09	**0.44**					
Long-term gov. bond rate	7.05	3.39	5.63	2.76	**0.49**	7.60	3.20	7.39	3.50	0.06
Short-term gov. bond rate	5.52	3.54	4.39	2.78	**0.38**	6.86	3.59	6.27	3.80	0.16
Policy rate	6.15	3.97	5.59	3.14	0.17	7.31	3.54	6.80	3.58	0.14
Macroeconomic Environment										
Real GDP per captia	0.01	0.01	0.00	0.01	0.15	0.01	0.01	0.01	0.01	0.00
Inflation	0.01	0.01	0.01	0.01	0.08	5.43	4.61	5.91	5.08	-0.10
House prices	0.02	0.02	0.02	0.03	0.13	8.29	8.48	8.62	10.09	-0.04
Unemployment rate	6.16	1.38	5.47	1.50	**0.46**	5.42	3.19	4.29	3.07	0.36
Consumer confidence	88.60	12.34	88.06	11.06	0.05					
Fiscal Variables										
Debt/GDP	43.32	11.36	47.70	12.71	**-0.35**					
News spending	8.33	66.99	6.19	69.98	0.03					
Defense spending	0.00	0.02	0.01	0.04	**-0.28**					
Government spending	0.00	0.02	0.01	0.03	**-0.24**	0.00	0.01	0.00	0.01	0.00
Global Variables										
Oil price index 1	0.01	0.10	0.02	0.10	-0.13					
Oil price index 2	0.00	0.03	0.01	0.05	-0.17					

| | Mertens and Ravn (2012), Unanticipated | | | | | Mertens and Ravn (2012), Aanticipated | | | | |
| | Tax Changes | | No Tax Changes | | diff/SD | Tax Changes | | No Tax Changes | | diff/SD |
Variable	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Financial Variables										
SP500	0.97	0.46	0.90	0.40	0.16	1.04	0.48	0.90	0.39	**0.34**
SPXD	0.05	0.07	0.02	0.07	**0.38**	0.02	0.06	0.02	0.08	-0.10
AAA cor. bond rate	6.96	3.26	6.76	2.87	0.07	8.13	3.24	6.62	2.83	**0.52**
BAA cor. bond rate	7.94	3.65	7.67	3.17	0.08	9.16	3.65	7.52	3.13	**0.51**
Long-term gov. bond rate	6.36	3.34	6.00	2.83	0.13	7.41	3.32	5.87	2.80	**0.53**
Short-term gov. bond rate	4.71	3.16	4.74	2.90	-0.01	5.92	3.70	4.59	2.79	**0.46**
Policy rate	5.35	3.55	5.78	3.33	-0.13	6.55	4.13	5.60	3.23	**0.28**
Macroeconomic Environment										
Real GDP per captia	0.00	0.01	0.01	0.01	-0.04	0.01	0.01	0.00	0.01	**0.36**
Inflation	0.01	0.01	0.01	0.01	0.08	0.01	0.01	0.01	0.01	0.03
House prices	0.02	0.02	0.02	0.02	0.13	0.02	0.02	0.02	0.02	**0.28**
Unemployment rate	6.03	1.27	5.55	1.54	**0.32**	6.06	1.45	5.55	1.51	**0.33**
Consumer confidence	87.41	11.04	88.39	11.43	-0.09	89.05	13.17	88.15	11.11	0.08
Fiscal Variables										
Debt/GDP	41.56	11.46	47.15	12.48	**-0.45**	42.91	10.92	47.02	12.64	**-0.33**
News spending	-8.62	47.02	5.43	51.72	**-0.27**	17.75	85.11	1.83	45.23	**0.31**
Defense spending	-0.01	0.02	0.01	0.04	**-0.33**	0.00	0.02	0.01	0.04	**-0.28**
Government spending	0.00	0.02	0.01	0.03	-0.17	0.00	0.02	0.01	0.03	**-0.35**
Global Variables										
Oil price index 1	0.00	0.03	0.01	0.04	**-0.30**	0.01	0.02	0.01	0.04	0.13
Oil price index 2	0.00	0.08	0.02	0.10	-0.24	0.02	0.10	0.02	0.10	0.01

Table A.4: p-value for H_0: Variable does not predict narrative shock

	RR(2010)	Cloyne (2013)
Financial variables		
Stock price index	.69	.38
BAA corp bond rate	**.05**	**.04**
Long-term gov bond rate	**.04**	**.01**
Short-term gov bond rate	**.09**	**.02**
Policy rate	.19	**.03**
Macroeconomic environment		
Real GDP per capita	.61	.75
House prices	.30	.89
Unemployment rate	**.00**	**.01**
Consumer confidence	.26	
Fiscal variables		
Debt/GDP	**.00**	
News spending	.20	
Defense spending	.96	
Government spending	.89	.98
Global		
Brent	.58	.53

9

References

Alesina, A. and S. Ardagna (2010). Large changes in fiscal policy: taxes versus spending. In *Tax Policy and the Economy, Volume 24*, pp. 35–68. The University of Chicago Press.

Alesina, A., C. Favero, and F. Giavazzi (2014). The output effect of fiscal consolidations. Technical report, mimeo.

Bahaj, S. A. (2014). Systemic sovereign risk: macroeconomic implications in the euro area. Technical report, mimeo.

Bluedorn, J. and D. Leigh (2011). Revisiting the twin deficits hypothesis: the effect of fiscal consolidation on the current account. *IMF Economic Review 59*(4), 582–602.

Cloyne, J. (2013). Discretionary tax changes and the macroeconomy: new narrative evidence from the united kingdom. *American Economic Review 103*(4), 1507–1528.

Dominguez, K. M. and M. D. Shapiro (2013). Forecasting the recovery from the great recession: Is this time different? *American Economic Review P&P 103*(3), 147–52.

Feyrer, J. and J. Shambaugh (2012). Global savings and global investment: The transmission of identified fiscal shocks. *American Economic Journal: Economic Policy 4*(2), 95–114.

Guajardo, J., D. Leigh, and A. Pescatori (2014). Expansionary austerity: International evidence. *Journal of the European Economic Association 12*(4), 949–968.

Hayo, B. and M. Uhl (2014). The macroeconomic effects of legislated tax changes in germany. *Oxford Economic Papers 66*(2), 397–418.

IMF (2010, September). *Recovery, Risk, and Rebalancing*, Chapter 3. World Economic Outlook.

Mertens, K. and M. O. Ravn (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy 4*(2), 145–81.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review 103*(4), 1212–1247.

Perotti, R. (2012). The effects of tax shocks on output: Not so large, but not small either. *American Economic Journal: Economic Policy 4*(2), 214–37.

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

Romer, C. D. and D. H. Romer (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review 100*, 763–801.

Romer, C. D. and D. H. Romer (2014). Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Changes, 1952-1991. NBER Working Papers 20087, National Bureau of Economic Research.

Recent Issues