Check for
updates

Too complex to digest? Federal tax bills and their processing in US financial markets

Hamza Bennani[1] · Matthias Neuenkirch[2]

Abstract

In this paper, we analyze whether the textual complexity of tax bills affects financial markets. Based on the Flesch-Kincaid grade level of the 32 tax bills identified by Romer (Am Econ Rev 100(3):763–801, 2010)in the period 1962–2003, we assess the relationship between tax bills' textual complexity and financial markets in various windows around the signing of a bill. Our results show a negative (positive) and significant relationship between the present value of tax bills and changes in the 10-year government bond yields (S &P 500 returns). The magnitude of this relationship increases over time, suggesting that market participants underreact at first and need a couple of days to digest the information contained in the tax bills. This delay can be explained by the textual characteristics of the bills in the case of the 10-year yields as a lower readability partly counteracts the negative relationship for up to three days after the signing of a tax bill. In the case of the stock market, we find similar evidence, but only for a part of the readability measures employed in this paper.

Keywords Complexity · Financial markets · Readability · Tax bills

JEL Classification G14 · H20 · H30

> *"We must …transform a system that's become an endless source of confusion and resentment into one that is clear, simple, and fair for all — a tax code that no longer runs roughshod over Main Street America but ensures your families and firms incentives and rewards for hard work and risk-taking in an American future of strong economic growth."*
> *Ronald Reagan, May 1985*

✉ Matthias Neuenkirch
  neuenkirch@uni-trier.de

[1] Nantes University, Nantes, France

[2] Department of Economics, Trier University and CESifo, 54286 Trier, Germany

🍀 Springer

1 Introduction

With more than 75,000 pages, US federal tax bills are three times larger today than when President Ronald Reagan complained about their complexity, even though the Internal Revenue Service revised tax schedules and instructions to improve readability during the past decades. Hence, despite the regulatory efforts made to improve the readability of tax bills (US Comptroller General, 1978), multiple evidences indicate that these are still low in readability (see, e.g., Bradbury et al., 2018). This can have sizable and detrimental effects on economic activity since legislative complexity can be seen as an indicator of a poor quality of government and a source of corruption (Rose-Ackerman, 2007).[1]

Persistent legislative complexity might be the result of the rational behavior of self-interested politicians aiming to appropriate public rent and to modify the income distribution in their favor (Tullock et al., 2002). Indeed, complex legislation makes the applicable law ambiguous and enables members of the legislation to take advantage of the asymmetric information with private agents (Acemoglu, 2006, 2013; Aidt, 2013). One channel through which this might be possible is the textual complexity of tax bills and the resulting underreaction of investors to (complex) information.

To date, however, the literature has considered the effect of tax complexity mainly through the lens of the number of tax rates, tax bases, tax payments, and special provisions. For example, Saez (2010) finds that taxpayers do not bunch at kinks if marginal incentives are fully understood and labor supply could be freely adjusted, while Baake et al. (2004) suggest that a simplified tax system—which does not distinguish consumption goods and work-related goods—provides higher incentives to work and is less redistributive. Tax complexity, however, has other dimensions, such as the readability of tax documents, which has not been studied thus far.[2] Indeed, textually complex financial documents, such as those related to tax bills, might limit the investors' ability to extract information and could lead to an (initial) underreaction. Following this line of thought, Abeler and Jaeger (2015) conduct a controlled laboratory experiment and find that people underreact to complex tax incentives, measured through the number of pages of legislation, the readability of the legislation, and the number of exemptions. Looking at consumers, Chetty et al. (2009) document that these underreact to changes in non-salient taxes. Both papers explain their findings with the cognitive costs that individual have to bear to digest all available information.

Against this background, we study the textual complexity of federal tax bills and its consequences on financial markets. We argue that textual complexity might affect financial market participants' behavior through at least two channels. First, tax bill

---

[1] Administration and compliance costs (IMF & OECD, 2019) or depressed firm birth rates (Bacher & Brülhart, 2013) are two of the reasons for these detrimental effects.

[2] Moreover, documentation requirements or administration and filing costs could also affect taxpayers. As an illustration, Warskett et al. (1998) consider administration costs and self-selection, that is, the ability for taxpayers to disguise themselves for the purpose of taxation, as important to explain the complexity of the tax structure.

complexity can be considered as a "barrier to entry" into markets due to private costs of acquiring information about new legislation (Kaplow, 1996). In this context, complexity reflects additional transaction costs that might prevent investors to access more data. This reduces their willingness to collect or process the information and, ultimately, affects their trading behavior (Bloomfield, 2008; Grossman & Stiglitz, 1980; Lundholm, 1991). This channel is referred to as the "*incomplete revelation hypothesis.*"

The second channel is based on the literature on human psychological and cognitive processes and examines the role of financial information in shaping investor decision-making processes (Sims, 2005). Specifically, given the cognitive limitations of investors, information complexity might affect their ability to fully incorporate such information into investment decisions. Put differently, textual complexity adds noise to messages and receivers are unable to fully internalize this information (Jin et al., 2022). As a consequence, investors might underreact to information that is more difficult to read, which results in an inertial reaction of asset prices to new information (see, for instance, Dellavigna and Pollet, 2009). This channel is referred as the "*limited attention hypothesis.*" In contrast, improved processing fluency associated with more readable disclosures enhances investors' confidence so that these can rely on the information in their decision-making and react more adequately to the disclosures (Loughran & McDonald, 2014).[3]

Based on these considerations, we hypothesize that financial market participants underreact when they face a complex tax bill. To test this hypothesis, we first collect the 32 tax bills identified by Romer and Romer (2010) (henceforth RR) in the period 1962–2003 from the congressional archives.[4] We rely on this particular dataset since it is commonly used in the literature to identify fiscal policy shocks and allows differentiating the tax bills according to their revenue effects, the nature of the changes, and their motivation. Second, to measure the textual complexity of tax bills, we rely on the Flesch-Kincaid (FK) grade level index to generate a readability measure based on education grade levels (Kincaid et al., 1975). The most important benefit of this measure is that it is based on objective elements of the underlying texts. Hence, despite the recent advancements in text analysis, the FK index remains

---

[3] One might argue that while individual investors might be cognitively constrained by the textual complexity of financial documents, the experts, such as institutional investors and other market professionals, are not. Behavioral research suggests, however, that capital market professionals also suffer from limited attention in the same way as retail investors and that they underreact to complex information (Hopkins, 1996; Hirst & Hopkins, 1998; Lee, 2012; Li, 2008)

It has to be noted that—although most studies find a positive and significant relationship between underreaction and complexity — others find evidence of an overreaction. In a life cycle experiment using Amazon Mechanical Turk, Lu (2022) finds that subjects exhibit an overreaction to wealth taxes by decreasing their savings. He explains that this overreaction might be driven by subjects' incorrect understanding of the effects of wealth taxes on their savings. Nevertheless, Lu (2022) admits that his results are sensitive to the sample choice and complexity of taxation might also attenuate individuals' overreaction. Daniel et al. (2020) suggest that it is the time horizon that determines investors' underreaction or overreaction to information. Specifically, while short-term mispricing is driven by an underreaction, the long-term one is due to overconfidence.

[4] The starting date is restricted by the availability of daily data on 10-year US government bonds and the end date coincides with the last bill in the RR dataset.

very efficient and is widely used in the social science literature.[5] Finally, we assess the relationship between the tax bills' textual complexity and financial markets' behavior by considering the changes in 10-year government bond yields and S &P 500 returns in windows from one day before the signing of a tax bill until up to four days after the signing of a bill by the President.[6]

Our results show that financial market participants react significantly to the characteristics of tax bills after their signing. Specifically, we find a negative (positive) and significant relationship between the present value of tax bills and changes in the 10-year bond yields (S &P 500 returns), suggesting that market participants actually appreciate an increase in taxes. The magnitude of this relationship increases over time, indicating that market participants underreact at first and need a couple of days to digest the information contained in the tax bills. This delay can be explained by the textual characteristics in the case of the 10-year bond yields as a lower readability (proxied by the years of education required to understand a bill) partly counteracts the negative relationship for up to three days after the signing of a tax bill, but not thereafter. In the case of the stock market, we find similar evidence, but only for a part of the readability measures employed in this paper. We also test if the results differ according to the nature of the tax change (endogenous or exogenous) and the phase of the business cycle during which the bill is signed. Lastly, we document the robustness of our results with respect to (i) the usage of abnormal changes in bond yields (abnormal stock returns) instead of their raw counterparts, (ii) other textual characteristics of the tax bills, such as length, textual uncertainty, and sentiment, (iii) the timing of the tax bills (using the date of the vote in the Congress instead of the Presidential signing), (iv) the electoral cycle, (v) parameter instability, and (vi) outliers. Hence, we provide robust empirical evidence for an underreaction of investors around the day a bill is signed, which is explained by the textual complexity of the tax bills.

This paper relates to at least three strands of the literature. The first one analyzes the effect of tax changes using a narrative identification approach and the same dataset (e.g., Romer & Romer, 2010; Perotti, 2012; Mertens & Ravn, 2012, 2013). The second strand examines the effect of the complexity of financial disclosures on investors' behavior (e.g., Miller, 2010; You & Zhang, 2009). The third branch analyzes the connection between financial markets and tax policy (e.g., Gaertner et al., 2020; Ardagna, 2009; Wagner et al., 2018).[7]

The remainder of this paper is organized as follows. Section 2 describes our approach to measure the textual complexity of tax bills. Section 3 presents the empirical methodology. Section 4 shows the results. Section 5 concludes.

---

[5] This is documented by Benoit et al. (2019) who find the FK index to be a crucial predictor of sophistication in political texts, whereas the introduction of various additional text features only marginally improves the predictive capacity.

[6] We explain the choice of the signing date over, for instance, the date of the final vote in the Senate in Sect. 3 below and also test for the robustness of this choice in Sect. 4.3.

[7] There are also several papers studying the clarity of central bank communication and its impact on financial markets (e.g., Jansen, 2011; Ehrmann & Talmi, 2020; Hayo et al., 2022).

2 Measuring the textual complexity of tax bills

We first collect the dataset of tax bills identified by RR from the congressional archives for the period from October 1962 to May 2003 (see Table 12 in Appendix A for a list of the bills). Hence, the Revenue Act of 1962 (signed on 16-Oct-62) marks the first event in our sample—given the availability of daily 10-year government bond yields in the FRED database—and the Jobs and Growth Tax Relief Reconciliation Act of 2003 (signed on 28-May-03) marks the final one. We obtain a set of 32 scanned PDF files that need to be converted into text files in order to be processed. We utilize the *Tesseract* package in *R* that relies on a powerful optical character recognition (OCR). OCR is the process of finding and recognizing text inside a scanned paper by using language-specific training data in the recognized words. Specifically, *Tesseract* converts each single PDF file into a PNG format and then scans for texts within the created images.[8]

As a second step, we rely on Schuck (1992)'s definition to measure the complexity of the tax bills.[9] Out of the four dimensions he identifies, we focus on the density and conjecture that the complexity of a tax bill can be objectively measured with the characteristics of its text. Indeed, social scientists usually proxy text complexity by its readability, that is, the easiness for an individual to read a text and to understand the information. Following this line of thought, past research (Flesch, 1948) has identified text characteristics, such as the length of words and sentences, as good predictors of readability.

To quantify the readability of tax bills and, thus, their complexity, several measures—such as the Flesch-Kincaid (FK) grade level (Kincaid et al., 1975)—exist, which can be interpreted as the number of years of education needed to sufficiently comprehend a text.[10] The FK index has been used in a variety of fields, spanning from medicine to psychology, also including many studies in economics (e.g., Jansen, 2011a) and political science (e.g., Schoonvelde et al., 2019). The FK index is calculated as follows:

$$0.39 \cdot (\#words/\#sentences) + 11.8 \cdot (\#syllables/\#words) - 15.59 \tag{1}$$

(# words /# sentences) reflects the number of words per sentence and (# syllables/# words) the number of syllables per word. The rationale underlying this index is that many words per sentence or many syllables per word in a text decrease (increase) its readability (complexity) and, therefore, require more years of education to be understood.

Figure 1 shows the FK index of the tax bills from 1962 until 2003. Most of the federal tax bills in our sample were implemented during the period between 1960

---

[8] Other than that, we did not do any pre-processing (e.g., removing punctuation or stemming) since we want to create a measure of complexity from a reader's perspective. More technically, the Flesch-Kincaid measure introduced below is based on the number of syllables in a word and the number of words in a sentence. Hence, manipulating the texts would distort this measure.

[9] The four characteristics to define legal complexity according to Schuck (1992) are density, technicality, differentiation, and indeterminacy.

[10] In Sect. 4.3, we use alternative readability measures as part of our robustness tests.

and 1985. There is no clear upward or downward trend in the complexity of tax bills over time.[11] Additional information about the present value of the tax bills (see Fig. 3 in Appendix A) does not reveal any substantial correlation between these two dimensions (see Table 14 in Appendix A). Lastly, it has to be noted that the tax bill from 15-Mar-1966 (Tax Adjustment Act of 1966) is clearly an outlier in terms of its textual complexity and will be excluded as part of our robustness tests in Sect. 4.3.

3 Empirical methodology

Our sample consists of the 32 legislative texts identified by RR for which we have financial market data at hand. We consider the signing of each bill as an event and are interested in the movement of financial markets around the events. Figure 4 shows the average daily changes in the 10-year government bond yields as well as the average daily returns of the S &P 500 returns two days before until four days after the signing of a tax bill. We do not observe a clear pattern in the data, but this should not be surprising. Tax bills have at least two dimensions that should matter for market participants.

The first one is the *present value* of taxes (as % of nominal GDP), which has been used in various studies to analyze the effect of tax shocks (see, e.g., Romer & Romer, 2010; Perotti, 2012; Mertens & Ravn, 2012, 2013). Our sample consists of 17 bills that lead to an increase in the present value of taxes and 15 tax cuts, with the Economic Recovery Tax Act of 1981 (signed on 13-Aug-81) being the largest (with a present value of −3.96% of GDP).[12] An increase in taxes should lead to lower refinancing costs of the government and—as a contractionary fiscal policy measure—should also dampen economic growth (at least in the short-run). Both channels indicate a negative relationship between the present value of a tax bill and bond yields after the signing of the tax bill. The effect on stock returns, however, is a priori unclear. An improvement in public finances should lead to a lower discount rate, but a higher tax burden should reduce (future) cash flows. Hence, the discounted present value of firms could either increase or decrease.

The second dimension is the *textual complexity* of tax bills. As indicated in Sect. 1, market participants might need some time to digest the information in the bill in order to update their information set and to adjust their trading behavior. Hence, we expect that the maximum effect of a tax bill on bond yields and stock returns materializes with a time lag due to an underreaction of the investors. In particular, we conjecture that a higher level of textual complexity contributes to this delay and we expect an opposing sign compared to the one of the present value. Accordingly, our two hypotheses are as follows:

Hypothesis 1 Tax increases (cuts) lead to a decline (an increase) in 10-year bond yields. The effect of tax changes on stock returns is ambiguous.

---

[11] Regressing the FK index on a trend parameter yields a positive, but insignificant trend estimate (p-value: 0.73).

[12] Fig. 3 in Appendix A shows the present value for each tax bill and Table 13 lists descriptive statistics.

Hypothesis 2 The maximum effect of a tax change can be found after a time lag. Higher textual complexity contributes to this transmission delay.

To test how tax bills are processed by financial market participants, we consider several windows around each event corresponding to the President's signing date of a bill. Indeed, even though the legislative process in the US starts long before a bill is signed by the President and the fiscal measures are publicly debated in the meantime, several papers find no evidence in favor of anticipation effects. For instance, Mertens and Ravn (2012) use a timing convention to distinguish between anticipated and unanticipated changes in taxes.[13] Their finding of no strong anticipation effects is in line with several studies showing that it is the implementation date of tax changes that mainly affects macroeconomic variables (Poterba, 1988; Parker, 1999; Souleles, 1999, 2002; Mertens & Ravn, 2010). More recently, Hayo and Mierzwa (2020) find a significant reaction of financial markets on the days changes in tax legislation are actually implemented. These findings might also be explained by the prevailing uncertainty until the signing date of a tax bill, which could be due to the relatively large number of vetoes made by US Presidents in the 1960 s and 1970 s (for instance, Richard M. Nixon made 26 vetoes while Gerald R. Ford made 48 vetoes).[14] This raises one crucial question for financial markets: Will the bill actually pass the final step of the legislative process?

Each window begins one day before the event. The end of a window can be either on the day after the signing or up to four days thereafter to capture the information processing. This yields a total of four event windows, which are summarized in Fig. 2. We calculate the changes in the 10-year bond yields as well as the growth rate in the S &P 500 index over each of these windows.

Next, we explain these changes in bond yields and stock returns using the following specification:

$$y_{e,w} = \alpha + \beta PV_e + \gamma FK_e + \epsilon_{e,w} \tag{2}$$

$y_{e,w}$ is either (i) the change in 10-year government bond yields or (ii) the growth rate in the S &P 500 index for event e during one of the windows w introduced in Fig. 2. α is a constant term, β and γ are slope parameters for the present value (PV_e) of a tax bill and its textual complexity (FK_e), and $\epsilon_{e,w}$ is the error term.[15] Eq. (2) is estimated using least squares with heteroskedasticity-consistent standard errors. Both explanatory variables are demeaned, so that the constant term represents the conditional average change in the 10-year bond yields or the conditional average growth rate of

---

[13] Unfortunately, their distinction leaves too few observations for a meaningful analysis in the context of this paper.

[14] A full list of vetoes can be found at: https://www.senate.gov/legislative/vetoes/vetoCounts.htm.

[15] As part of our robustness tests, we replace the "raw" changes/returns in Eq. (2) with the "abnormal" changes/returns over each of the four event windows. For that purpose, we calculate the average daily changes of the 10-year bond yields (S &P 500 returns) over a window from 180 days to 30 days before each event. We subtract these daily averages from the actual daily changes in bond yields (daily stock returns) to calculate the abnormal changes (returns) over the two- to five-day windows. The results of this robustness test can be found in Sect. 4.3.
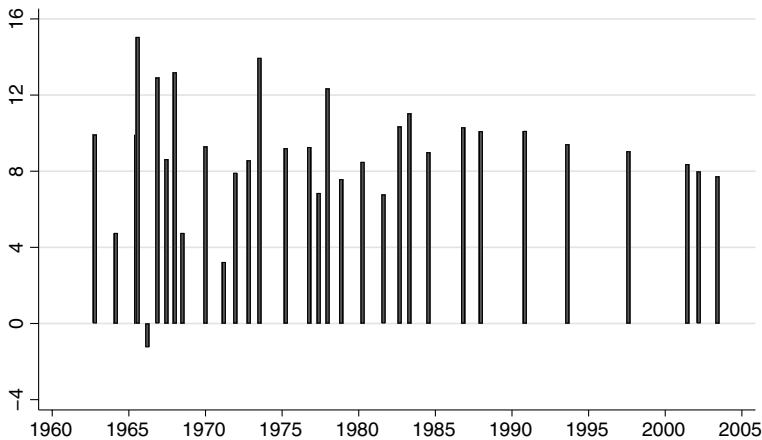
Fig. 1 The Flesch-Kincaid Grading Level of Tax Bills. *Notes:* Figure shows the textual complexity of the tax bills (as measured by the FK grade level). The corresponding tax bills can be found in Table 12

the S &P 500 for an event window. It has to be noted that we also considered including an interaction term between the two slope parameters. This term, however, generates no additional insights and is excluded from the analysis.[16]

Our analysis is based on a couple of assumptions. First, we assume that there are no confounding factors. This assumption can be easily justified as the signing of tax bills is the result of a (lengthy) political process and the actual signing day should not be systematically related to, for instance, monetary policy decisions or the publication of important macroeconomic news.[17] Second, it is typically the unexpected component of an announcement that should actually matter for financial market participants. In an ideal world, we would extract such a news component from market expectations as it is done in the literature on macroeconomic news or monetary policy decisions and their impact on financial markets. However, in the absence of a series for the expected present value (textual complexity), we consider the actually observed value as second-best proxy.

Finally, we also considered studying the response of the volatility of both financial series around the signing of a tax bill. For that purpose, we extract the conditional standard deviation of the change in the 10-year bond yields and the S &P 500 returns with the help of an AR(1)-GARCH(1,1) model and t-distributed errors. The results (available on request) indicate no systematic relationship of the conditional standard deviations with the present value of tax bills or their textual complexity in the four event windows.

---

[16] All omitted results are available on request.

[17] As part of our robustness tests, we remove those bills from the analysis that were signed close to US monetary policy decisions (see Sect. 4.3).
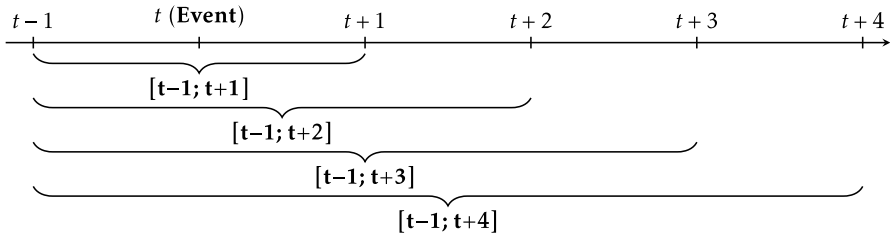
Fig. 2 Event Windows Around the Signing of a Tax Bill

4 Results

4.1 Baseline results

Tables 1 and 2 show the results of Eq. (2) when using the 10-year bond yields and the S &P 500 returns as the dependent variable, respectively.

We find a negative and significant relationship between the present value of tax bills and the change in the 10-year bond yields (in line with H1). In line with H2, this relationship increases over time, suggesting that financial market participants underreact at first and need a couple of days to digest the information contained in the tax bills. In terms of economic magnitude, we find that a one percentage point (pp) increase in the present value of tax bills lowers the 10-year bond yield by 3.1 basis points (bps) on the day after the bill is signed (as compared to $t-1$), while the response increases to -6.8 bps three and -9.3 bps four days after the signing. Stock market returns, on the other hand, increase significantly by 63.5 bps two days after a bill is signed, and the response amounts to 84.9 bps for the event window *[t−1; t+4]*. Hence, we find evidence that tax consolidation is actually appreciated by financial markets.

The complexity of tax bills, measured by the FK grading level, provides additional insights about market participants' behavior when confronted with difficult-to-read texts. We observe that a lower readability (i.e., higher complexity) is positively associated with the changes in the 10-year bond yields, whereas the effect of textual complexity on bond yields slightly increases over time (up to 1.2 bps after two days), it is no longer significant four days after the signing. Hence, in line with H2, higher textual complexity contributes to the delay in the processing of information by market participants. This result, however, is not replicated for the stock market as the coefficients for complexity are insignificant at the 10% level.

4.2 Subsample analysis

One of the most important contributions of the narrative approach by RR is to demonstrate that exogenous fiscal policy shocks have a larger and more significant effect on output as compared to broader measures of tax changes. Additional studies, such as Auerbach and Gorodnichenko (2012) and Fazzari et al. (2015), find that the effect of exogenous tax changes on output varies over the business cycle and that it is larger during economic slack. Against this background, we test whether the effect

of the present value and the textual complexity of the tax bills on financial markets' behavior is conditional on the nature of tax bills or the business cycle phase during which they are signed. Specifically, we distinguish between (i) endogenous and (ii) exogenous tax bills relying on the classification of RR and between tax bills implemented during (iii) an expansion and (iv) a recession relying on the NBER business cycle chronology.[18] We focus on [$t-1$; $t+2$] in this subsample analysis as the baseline results show the largest effects for textual complexity in this event window. Tables 3 and 4 show the results.

When considering the changes in the 10-year bond yields, the results for the present value and textual complexity are qualitatively similar for exogenous (albeit insignificant at the 10% level for the present value) and endogenous tax changes as well as for bills signed during an expansion. The only difference can be found for bills signed during a recession. However, the latter results have to be taken with a grain of salt due to the low number of observations in this subsample. For the S &P 500 returns, we find that the textual complexity of exogenous tax changes and those signed during an expansion significantly matter in a way consistent with H2 as there is an underreaction to more complex bills. Other than that, we find qualitatively similar results for the present value of exogenous tax changes, those signed during an expansion, and for endogenous tax changes (albeit not significant at the 10% level).

4.3 Robustness tests

4.3.1 "Abnormal" changes in bond yields and stock returns

First, we replace the "raw" changes/returns in Eq. (2) with the "abnormal" changes/returns over each of the four event windows. For that purpose, we calculate the average daily changes of the 10-year bond yields (S &P 500 returns) over a window from 180 days to 30 days before each event. We subtract these daily averages from the actual daily changes in bond yields (daily stock returns) to calculate the abnormal changes (returns) over the two- to five-day windows. The results of this robustness test are shown in Tables 5 and 6.

The coefficients are qualitatively and quantitatively almost unchanged when compared to the baseline results in Tables 1 and 2, in particular for the bond market. Hence, our results are robust to using the "abnormal" changes/returns instead of the "raw" changes/returns. This is also the reason why we stick to the usage of the non-transformed left-hand side variables through the remainder of this paper.

Different readability measures To test if our results are robust to the readability measure used in the specification, we use different approaches gaging text clarity: (i) the Flesch reading ease (FRE) index, (ii) the Gunning Fog (GF) index, the (iii) Simple Measure of Gobbledygook (SMOG) grade, and (iv) the new Dale-Chall (DC) readability formula. The FRE, the GF, and the SMOG indexes are similar to the FK grading level as these indicate the easiness to read a text and the years of formal

---

[18] It has to be noted that some of the tax bills contain a part exogenous to the business cycle and an endogenous part.

Table 1 Change in 10-year bond yields

Event window	$[t-1; t+1]$	$[t-1; t+2]$	$[t-1; t+3]$	$[t-1; t+4]$
Present value	−0.031*	−0.037**	−0.068***	−0.093**
	(0.015)	(0.016)	(0.015)	(0.040)
	[0.055]	[0.030]	[0.000]	[0.028]
FK grading level	0.007**	0.012***	0.011***	0.004
	(0.004)	(0.004)	(0.004)	(0.009)
	[0.050]	[0.004]	[0.005]	[0.615]
Constant	−0.023	−0.019	−0.037*	−0.042
	(0.019)	(0.019)	(0.021)	(0.029)
	[0.239]	[0.315]	[0.094]	[0.152]
R^2	0.089	0.154	0.253	0.253

Table shows results for a least squares estimation of Eq. (2) and different event windows with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

Table 2 S &P 500 Returns

Event window	$[t-1; t+1]$	$[t-1; t+2]$	$[t-1; t+3]$	$[t-1; t+4]$
Present value	0.234	0.635**	0.792***	0.849***
	(0.185)	(0.232)	(0.250)	(0.236)
	[0.218]	[0.010]	[0.004]	[0.001]
FK grading level	0.030	−0.101	−0.086	−0.067
	(0.088)	(0.067)	(0.069)	(0.068)
	[0.739]	[0.142]	[0.223]	[0.334]
Constant	−0.207	−0.120	0.302	0.320
	(0.233)	(0.280)	(0.341)	(0.353)
	[0.383]	[0.673]	[0.383]	[0.372]
R^2	0.047	0.145	0.147	0.157

Table shows results for a least squares estimation of Eq. (2) and different event windows with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

education a person needs to understand the text on the first reading. While the Flesch reading-ease (FRE) is computed with the same formula as the FK grading level but with different values, the GF index, and the SMOG grade are calculated using complex words (polysyllables), that is, words consisting of two or more syllables.[19] The

---

[19] The alternative measures are calculated as follows:

FRE: $206.835 - 1.015 \cdot (\#words/\#sentences) - 84.6 \cdot (\#syllables/\#words)$;

GF: $0.4 \cdot [(\#words/\#sentences) + 100 \cdot (\#complex\,words/\#words)]$;

SMOG: $1.043 \cdot \sqrt{\#polysyllables \cdot (30/\#sentences)} + 3.1291.$

Table 3 Change in 10-year bond yields: subsample analysis

	Full	Exogenous	Endogenous	Expansion	Recession
Present value	−0.037**	−0.028	−0.069**	−0.054***	−0.191**
	(0.016)	(0.021)	(0.026)	(0.014)	(0.044)
	[0.030]	[0.195]	[0.025]	[0.001]	[0.023]
FK grading level	0.012***	0.012**	0.012*	0.011***	0.271**
	(0.004)	(0.006)	(0.005)	(0.003)	(0.072)
	[0.004]	[0.039]	[0.057]	[0.003]	[0.033]
Constant	−0.019	−0.013	−0.013	−0.006	−0.163**
	(0.019)	(0.028)	(0.027)	(0.012)	(0.050)
	[0.315]	[0.655]	[0.632]	[0.637]	[0.047]
Observations	32	23	12	26	6
R^2	0.154	0.088	0.430	0.366	0.836

Table shows results for a least squares estimation of Eq. (2) and the event window [$t-1$; $t+2$] for different subsamples with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Endogenous/Exogenous: Classification of tax changes according to Romer and Romer (2010); Expansion/Recession: According to NBER definition

Table 4 S &P 500 returns: subsample analysis

	Full	Exogenous	Endogenous	Expansion	Recession
Present value	0.635**	0.453**	1.226	1.008*	0.210
	(0.232)	(0.190)	(0.682)	(0.500)	(0.599)
	[0.010]	[0.027]	[0.106]	[0.055]	[0.750]
FK grading level	−0.101	−0.191**	−0.005	−0.125*	0.277
	(0.067)	(0.079)	(0.095)	(0.065)	(0.852)
	[0.142]	[0.025]	[0.962]	[0.069]	[0.766]
Constant	−0.120	−0.350	−0.137	−0.241	−0.079
	(0.280)	(0.367)	(0.495)	(0.362)	(0.721)
	[0.673]	[0.351]	[0.789]	[0.512]	[0.920]
Observations	32	23	12	26	6
R^2	0.145	0.112	0.379	0.169	0.336

Table shows results for a least squares estimation of Eq. (2) and the event window [$t-1$; $t+2$] for different subsamples with heteroskedasticity-consistent standard errors in parentheses and *p*-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Endogenous/Exogenous: Classification of tax changes according to Romer and Romer (2010); Expansion/Recession: According to NBER definition

new DC index uses a count of "hard" words and sentence length to calculate the US grade level of a text.[20] Since all five measures (FK, FRE, GF, SMOG, and DC) show a substantial correlation (see also Table 14 in the Appendix), we additionally use the

---

[20] These "hard" words do not appear in a specially designed list of common words familiar to most 4th grade students. The new DC index is computed as follows:

$[0.1579 \cdot (\text{\#hardwords/\#words}) + 0.0496 \cdot (\text{\#words/\#sentences})] + 3.6365.$

Table 5 "Abnormal" changes in 10-year bond yields

Event window	[$t-1$; $t+1$]	[$t-1$; $t+2$]	[$t-1$; $t+3$]	[$t-1$; $t+4$]
Present Value	−0.031*	−0.038**	−0.069***	−0.094**
	(0.017)	(0.019)	(0.016)	(0.040)
	[0.074]	[0.049]	[0.000]	[0.024]
FK grading level	0.007*	0.011***	0.011**	0.004
	(0.004)	(0.003)	(0.004)	(0.009)
	[0.053]	[0.003]	[0.015]	[0.694]
Constant	−0.027	−0.025	−0.045*	−0.052
	(0.021)	(0.021)	(0.024)	(0.033)
	[0.210]	[0.244]	[0.078]	[0.126]
R^2	0.076	0.127	0.208	0.212

Table shows results for a least squares estimation of Eq. (2) and different event windows with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. The left-hand side variables are the "abnormal" changes in the 10-year bond yields over the four event windows. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

Table 6 "Abnormal" S &P 500 returns

Event Window	[$t-1$; $t+1$]	[$t-1$; $t+2$]	[$t-1$; $t+3$]	[$t-1$; $t+4$]
Present value	0.235	0.637**	0.795***	0.853***
	(0.191)	(0.243)	(0.267)	(0.224)
	[0.227]	[0.014]	[0.006]	[0.001]
FK grading level	0.052	−0.068	−0.042	−0.012
	(0.093)	(0.075)	(0.078)	(0.074)
	[0.585]	[0.374]	[0.589]	[0.873]
Constant	−0.214	−0.130	0.287	0.302
	(0.238)	(0.286)	(0.361)	(0.370)
	[0.377]	[0.652]	[0.433]	[0.420]
R^2	0.061	0.137	0.137	0.155

Table shows results for a least squares estimation of Eq. (2) and different event windows with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. The left-hand side variables are the "abnormal" S &P 500 returns over the four event windows. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

(standardized) first principal component (PC1) of these as explanatory variable in the robustness test.[21] We replace the FK grading level with the different readability measures and re-estimate Eq. (2) in the event window [$t-1$; $t+2$]. It has to be noted that the FRE measure is defined as actual reading *ease* measure (and not a

---

[21] The first principal component explains 93.7% of the common variation in the five readability measures.

complexity measure). Hence, we multiply its coefficient with −1 to facilitate a comparison of the results across readability measures. Tables 7 and 8 show the results.

The results for the 10-year bond yields are robust, irrespective of the readability measure used in the estimation. Higher textual complexity leads to an underreaction of bond returns. In the case of stock returns, we now find some support for a significant effect of the complexity of tax bills for four (FRE, SMOG, DC, and PC1) out of the six complexity as these partly counteract the positive effects of the present value on stock returns. Hence, when replacing the FK index, we also find results that are consistent with H2 as there is an initial underreaction of the stock market, which can be (partly) explained by the textual complexity of the tax bills.

4.3.2 Length, Uncertainty and Sentiment

In a third step, we test if our results are not confounded by other textual characteristics, such as the length of the text, textual uncertainty, and the sentiment conveyed in a tax bill. Tan et al. (2014) suggest that low readability can magnify the effect of language sentiment if the former obfuscates readers' understanding. Hence, we extend our baseline specification and control for the number of words in tax bills (in logs), the textual uncertainty of tax bills, and the sentiment of tax bills. The measures for textual uncertainty (the share of words with an uncertain connotation) and sentiment (the share of positively connotated words minus the share of negatively connotated words) are obtained using the dictionary of Loughran and McDonald (2011). Tables 9 and 10 show the results.

The results for the 10-year bond yields are robust, irrespective of any additionally included covariate. If at all, the coefficient for the present value loses significance once sentiment is included into Eq. (2). In the case of the S &P 500 returns, we find some variation in the size of the coefficient for the present value (between 50.6 bps and 74.1 bps). The smallest coefficient can be explained by the inclusion of the sentiment variable, which positively affects stock returns as a 1pp higher sentiment leads to an increase of 124.3 bps in the S &P 500 returns. Lastly, textual complexity remains insignificant on the stock market.

4.3.3 Congressional vote vs. presidential signing

Every bill in the US has to pass both chambers of the Congress before the President has to sign or veto the enrolled bill. Hence, there is some uncertainty in the electoral process even after the final vote of the Senate. Still, market participants might already react to the present value and the complexity of a bill after it has passed the Congress. Accordingly, we collect the dates of the final vote in the Senate and test if market participants react at that time. Table 11 sets out the results with column 'President' replicating the baseline results using the restricted sample.[22] A couple of things are worth highlighting. First, the R^2 is higher in the three-day windows around the singing of the bill by the President. In addition, we observe larger (absolute) coefficients with

---

[22] For one bill (Tax Reform Act of 1969), the date of the final vote in the Senate is not available on the Congress website. The Revenue Act of 1962 passed the senate in 1961; however, daily data for the 10-year bond yields are only available since 1962. Hence, the number of observations shrinks to 30 as compared to 32 in the previous estimations.

a higher level of significance in the columns 'President' as compared to the columns 'Congress.' Hence, this reaffirms our choice of using the signing day by the President as endpoint for the electoral process as the weaker results for the Congress vote might indeed reflect the remaining uncertainty in the process.

4.3.4 Electoral cycle

Three of the bills in our sample (8-Nov-1966, 6-Nov-1978, and 5-Nov-1990) were signed on an election day or on the day before. Hence, the results could be confounded by the financial markets' reaction to these three midterm elections. Alternatively, market participants could react differently to bills signed close before an election. As a consequence, we exclude the three bills and re-estimate Eq. (2) without these in a first step. Furthermore, four additional bills (16-Oct-1962, 30-Oct-1972, 4-Oct-1976, and 22-Oct-1986) were signed in the month before an Presidential or midterm election. Accordingly, we exclude all seven bills to account for the potentially confounding electoral cycle in a second step. The results[23] remain robust for the bond market with a minor shrinkage in the significance of the coefficient for the present value. On the stock market, the coefficient for the present values becomes smaller, but remains quantitatively robust. Hence, we are confident that our results are not confounded by the electoral cycle.

4.3.5 Parameter stability

Next, we test for the stability of the estimated parameters over time. With the emergence of the Internet, information might be spread quicker and computer technology might also be helpful in digesting the complexity of the tax bills. To test for potential breaks in the parameters, we re-estimate Eq. (2) recursively using an expanding window with a minimum size of ten tax bills. The results can be found in Fig. 5 in Appendix B. The coefficients for the present value and textual complexity are significant on the bond market for all windows considered in the recursive estimations. If at all, the estimate for the present value decreases over time (in absolute terms), whereas the estimate for textual complexity increases. In the case of the stock market, the estimate for the present value does not show a distinct tendency over time with respect to its size; however, it becomes statistically significant only if a certain number of observations is included in the analysis. In general, we do not find evidence of a distinct structural break in the data.

4.3.6 US monetary policy decisions

Three of the bills in our sample (9-Jul-73, 20-Dec-77, and 18-Jul-84) are signed on the day of US monetary policy decisions or the day thereafter. Removing these from the analysis leads to a virtually unchanged coefficient for textual complexity on the bond market. On the stock market, the coefficient becomes slightly larger (its absolute value) and significant at the 10% level. Hence, we find no evidence that decisions by the Federal Open Market Committee confound our results.

---

[23] To conserve space, we do not report these results in detail. All omitted results are available on request.

Table 7 Change in 10-year bond yields: different readability measures

	FK	FRE	GF	SMOG	DC	PC1
Present value	−0.037**	−0.034**	−0.038**	−0.037**	−0.026*	−0.035**
	(0.016)	(0.015)	(0.017)	(0.016)	(0.015)	(0.016)
	[0.030]	[0.034]	[0.030]	[0.028]	[0.096]	[0.036]
Readability	0.012***	0.003***	0.011***	0.019***	0.028**	0.035***
	(0.004)	(0.001)	(0.004)	(0.006)	(0.013)	(0.012)
	[0.004]	[0.007]	[0.004]	[0.005]	[0.039]	[0.006]
Constant	−0.019	−0.019	−0.019	−0.019	−0.019	−0.019
	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)
	[0.315]	[0.314]	[0.317]	[0.316]	[0.329]	[0.317]
R^2	0.154	0.159	0.148	0.151	0.104	0.147

Table shows results for a least squares estimation of Eq. (2) and the event window [**t−1**; **t+2**] for different readability measures with heteroskedasticity-consistent standard errors in parentheses and p-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. *FK* Flesch-Kincaid Grading Level, *FRE* Flesch Reading Ease multiplied with −1 to facilitate a comparison of results across readability measures, *GF* Gunning Fog Index, *SMOG* SMOG Index, *DC* Dale-Chall Readability, *PC1* First principal component of the five readability measures. Number of observations: 32

Table 8 S &P 500 returns: different readability measures

	FK	FRE	GF	SMOG	DC	PC1
Present value	0.635**	0.622***	0.639**	0.645***	0.543**	0.628***
	(0.232)	(0.220)	(0.235)	(0.230)	(0.207)	(0.224)
	[0.010]	[0.008]	[0.011]	[0.009]	[0.014]	[0.009]
Readability	−0.101	−0.033*	−0.095	−0.180*	−0.486**	−0.364*
	(0.067)	(0.018)	(0.066)	(0.105)	(0.217)	(0.200)
	[0.142]	[0.081]	[0.164]	[0.099]	[0.033]	[0.079]
Constant	−0.120	−0.120	−0.120	−0.120	−0.120	−0.120
	(0.280)	(0.277)	(0.281)	(0.279)	(0.275)	(0.278)
	[0.673]	[0.669]	[0.673]	[0.671]	[0.667]	[0.670]
R^2	0.145	0.162	0.141	0.152	0.177	0.157

Table shows results for a least squares estimation of Eq. (2) and the event window [**t−1**; **t+2**] for different readability measures with heteroskedasticity-consistent standard errors in parentheses and *p*-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. *FK* Flesch-Kincaid Grading Level, *FRE* Flesch Reading Ease multiplied with −1 to facilitate a comparison of results across readability measures, *GF* Gunning Fog Index, *SMOG* SMOG Index, *DC* Dale-Chall Readability, *PC1* First principal component of the five readability measures. Number of observations: 32

4.3.7 Outlier analysis

As mentioned in Sect. 2, the Tax Adjustment Act of 1966 (Text No 5 in Table 12) is an outlier. It contains a lot of numbers and short half-sentences and, therefore, the FK measure fails to grade this text in an "appropriate" way. Accordingly, we test whether our results are robust when excluding this text. In the case of the 10-year

Table 9 Change in 10-year bond yields: length and sentiment

	Baseline	Length	Text. Unc.	Sentiment
Present value	−0.037**	−0.036**	−0.038**	−0.033*
	(0.016)	(0.017)	(0.018)	(0.018)
FK grading level	0.012***	0.011**	0.011***	0.011***
	(0.004)	(0.004)	(0.004)	(0.004)
Log(Words)		0.005		
		(0.011)		
Textual uncertainty			0.025	
			(0.163)	
Sentiment				−0.039
				(0.042)
Constant	−0.019	−0.019	−0.019	−0.019
	(0.019)	(0.019)	(0.019)	(0.019)
R^2	0.154	0.158	0.154	0.173

Table shows results for a least squares estimation of Eq. (2) and the event window [**t−1**; **t+2**], controlling for the length of tax bills (number of words, in logs), the textual uncertainty of tax bills, and the sentiment of tax bills with heteroskedasticity-consistent standard errors in parentheses. The measures for textual uncertainty (the share of words with an uncertain connotation) and sentiment (the share of positively connotated words minus the share of negatively connotated words) are obtained using the dictionary of Loughran and McDonald (2011). All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

bond yields, the coefficients are slightly larger when omitting the "questionable" text; for the stock returns, the results are qualitatively unchanged despite some slightly smaller coefficients without the outlier.

Our results also remain robust if we exclude the Economic Recovery Tax Act of 1981 (signed on 13-Aug-81, Text No 21 in Table 12) from the analysis. This tax cut (the largest one in our sample as mentioned in Sect. 3) was at the center of Ronald Reagan's Presidential campaign and, therefore, largely anticipated by market participants. In this case, the coefficient for the present value is slightly larger for bond yields and stocks returns without the outlier, whereas the coefficients for textual complexity remain virtually unchanged.

5 Conclusions

In this paper, we analyze whether the textual complexity of tax bills affects financial market participants' behavior. We first collect the 32 tax bills identified by Romer and Romer (2010) in the period 1962–2003 from the congressional archives. Second, to measure the textual complexity of tax bills, we mainly rely on the Flesch-Kincaid grade level index to generate a readability measure based on education grade levels. Lastly, we assess the relationship between tax bills' textual complexity and financial markets' behavior by considering the changes in 10-year government bond yields and S &P 500 returns in windows from one day before the signing of a tax bill until up to four days after the signing.

Table 10 S &P 500 returns: length and sentiment

	Baseline	Length	Text. Unc.	Sentiment
Present value	0.635**	0.581**	0.741***	0.506**
	(0.232)	(0.244)	(0.263)	(0.217)
FK grading level	−0.101	−0.057	−0.035	−0.085
	(0.067)	(0.071)	(0.080)	(0.075)
Log(Words)		−0.240		
		(0.166)		
Textual uncertainty			−4.301	
			(3.057)	
Sentiment				1.243*
				(0.677)
Constant	−0.120	−0.120	−0.120	−0.120
	(0.280)	(0.278)	(0.274)	(0.270)
R^2	0.145	0.187	0.212	0.234

Table shows results for a least squares estimation of Eq. (2) and the event window [**t−1**; **t+2**], controlling for the length of tax bills (number of words, in logs), the textual uncertainty of tax bills, and the sentiment of tax bills with heteroskedasticity-consistent standard errors in parentheses. The measures for textual uncertainty (the share of words with an uncertain connotation) and sentiment (the share of positively connotated words minus the share of negatively connotated words) are obtained using the dictionary of Loughran and McDonald (2011). All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 32

Table 11 Congressional vote vs. presidential signing

	Change in 10-Year Yields		S &P 500 Returns	
	Congress	President	Congress	President
Present value	−0.026*	−0.035**	0.539**	0.682***
	(0.015)	(0.016)	(0.241)	(0.240)
	[0.089]	[0.036]	[0.033]	[0.008]
FK grading level	0.004	0.012***	0.045	−0.098
	(0.004)	(0.004)	(0.079)	(0.064)
	[0.303]	[0.005]	[0.573]	[0.138]
Constant	−0.012	−0.022	0.190	−0.042
	(0.016)	(0.020)	(0.330)	(0.273)
	[0.470]	[0.278]	[0.571]	[0.879]
R^2	0.089	0.140	0.113	0.179

Notes: Table shows results for a least squares estimation of Eq. (2) and the event window [**t−1**; **t+2**] around the final vote in the Congress and the signing of the bill by the President with heteroskedasticity-consistent standard errors in parentheses and *p*-values in brackets. All explanatory variables are demeaned. ***/**/* indicate significance at the 1%/5%/10% level. Number of observations: 30

Our results show that financial market participants react significantly to the characteristics of tax bills after their signing. Specifically, we find a negative (positive) and significant relationship between the present value of tax bills and changes in the 10-year bond yields (S &P 500 returns), suggesting that market participants actually appreciate an increase in taxes. The magnitude of this relationship increases over time, indicating that market participants underreact at first and need a couple of days to digest the information contained in the tax bills. This delay can be explained by the textual characteristics in the case of the 10-year bond yields as a lower readability (proxied by the years of education required to understand a bill) partly counteracts the negative relationship for up to three days after the signing of a tax bill, but not thereafter. In the case of the stock market, we find similar evidence, but only for a part of the readability measures employed in this paper. We also test if our results differ according to the nature of the tax change (endogenous or exogenous) and the phase of the business cycle. Lastly, we document the robustness of our results with respect to (i) the usage of abnormal changes in bond yields (abnormal stock returns) instead of their raw counterparts, (ii) other textual characteristics of the tax bills, such as length, textual uncertainty, and sentiment, (iii) the timing of the tax bills (using the date of the vote in the Congress instead of the Presidential signing), (iv) the electoral cycle, (v) parameter instability, and (vi) outliers.

This paper considers changes in government bond yields and the overall stock market. Indeed, there might be differences in the response to complex tax bills of, say, accounting firms and tech companies. Hence, an interesting avenue for future research would be to analyze the reaction of different industries or firms.

Appendix A: Background on dataset

See Tables 12, 13 and 14 and Figs. 4 and 5.

Table 12 Legislative texts

No	Date	Legislative text
1	16-Oct-62	Revenue Act of 1962
2	26-Feb-64	Revenue Act of 1964
3	21-Jun-65	Excise Tax Reduction Act of 1965
4	30-Jul-65	Social Security Amendments of 1965
5	15-Mar-66	Tax Adjustment Act of 1966
6	08-Nov-66	Public Law 89–800 (Suspension of Investment Tax Credit)
7	13-Jun-67	Public Law 90–26 (Restoration of the Investment Tax Credit)
8	02-Jan-68	Social Security Amendments of 1967
9	28-Jun-68	Revenue and Expenditure Control Act of 1968
10	30-Dec-69	Tax Reform Act of 1969
11	17-Mar-71	1971 Changes to Social Security
12	10-Dec-71	Revenue Act of 1971
13	30-Oct-72	Social Security Amendments of 1972
14	09-Jul-73	1973 Changes to Social Security
15	29-Mar-75	Tax Reduction Act of 1975
16	04-Oct-76	Tax Reform Act of 1976
17	23-May-77	Tax Reduction and Simplification Act of 1977
18	20-Dec-77	Social Security Amendments of 1977
19	06-Nov-78	Revenue Act of 1978
20	02-Apr-80	Crude Oil Windfall Profit Tax Act of 1980
21	13-Aug-81	Economic Recovery Tax Act of 1981
22	03-Sep-82	Tax Equity and Fiscal Responsibility Act of 1982
23	20-Apr-83	Social Security Amendments of 1983
24	18-Jul-84	Deficit Reduction Act of 1984
25	22-Oct-86	Tax Reform Act of 1986
26	22-Dec-87	Omnibus Budget Reconciliation Act of 1987
27	05-Nov-90	Omnibus Budget Reconciliation Act of 1990
28	10-Aug-93	Omnibus Budget Reconciliation Act of 1993
29	05-Aug-97	Balanced Budget Act of 1997 / Taxpayer Relief Act of 1997
30	07-Jun-01	Economic Growth and Tax Relief Reconciliation Act of 2001
31	09-Mar-02	Job Creation and Worker Assistance Act of 2002
32	28-May-03	Jobs and Growth Tax Relief Reconciliation Act of 2003

Table shows all legislative texts covered in the analysis. Source: Romer and Romer (2009)

Table 13 Descriptive statistics

	Mean	Std. Dev	Min	Max
Present value (% of Nominal GDP)	−0.09	1.03	−3.96	1.15
Flesch-kincaid grading level	8.90	3.15	−1.26	15.06
Flesch reading ease	65.45	11.44	49.44	107.84
Gunning fog index	12.16	3.19	2.24	18.67
SMOG index	11.42	1.95	4.68	14.67
Dale-Chall readability	10.30	0.88	6.55	11.06

Table shows descriptive statistics for the present value of the tax bills and their complexity

Table 14 Bivariate correlations

	PV	FK	FRE	GF	SMOG	DC
Present value	1					
FK grading level	0.31	1				
Flesch reading ease	−0.23	−0.97	1			
Gunning fog index	0.34	1.00	−0.95	1		
SMOG index	0.31	0.99	−0.98	0.99	1	
Dale-Chall readability	0.01	0.80	−0.90	0.77	0.83	1

Notes: Table shows bivariate correlations for the present value of the tax bills and their complexity. PV: present value (% of nominal GDP); FK: Flesch-Kincaid Grading Level; FRE: Flesch Reading Ease; GF: Gunning Fog Index; SMOG: SMOG Index; DC: Dale-Chall Readability
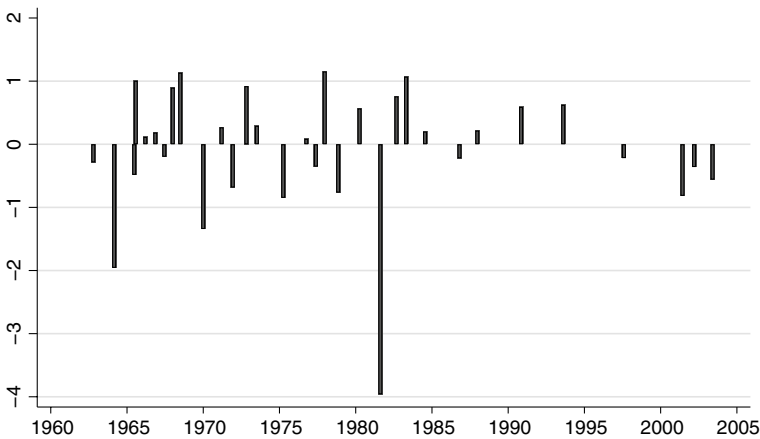
Fig. 3 Present value of tax bills. *Notes:* Figure shows the present value of the tax bills (as % of nominal GDP). The corresponding tax bills can be found in Table 12
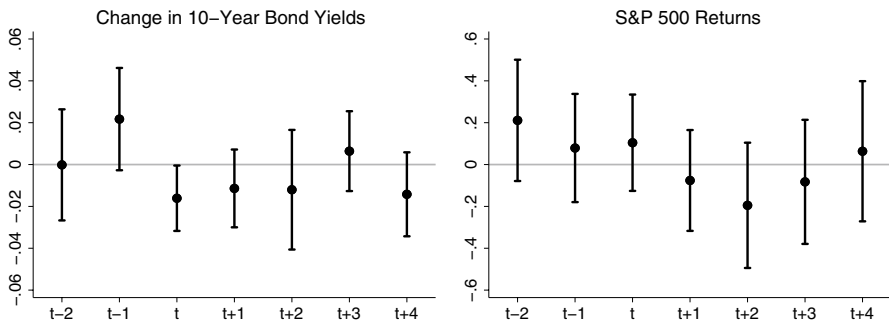
Fig. 4 Financial market data around the signing of tax bills. *Notes:* Figure shows the average daily change in the 10-year government bond yields as well as the average daily returns of the S &P 500 returns two days before until four days after the signing of a tax bill. 95% confidence intervals are indicated by whiskers. The average changes (returns) are obtained by a least squares regression (with robust standard errors) of the respective time series on a constant term and dummies for $t − 2$, $t − 1$, $\cdots$, and $t + 4$

Appendix B: Results of recursive estimations

See Fig. 5
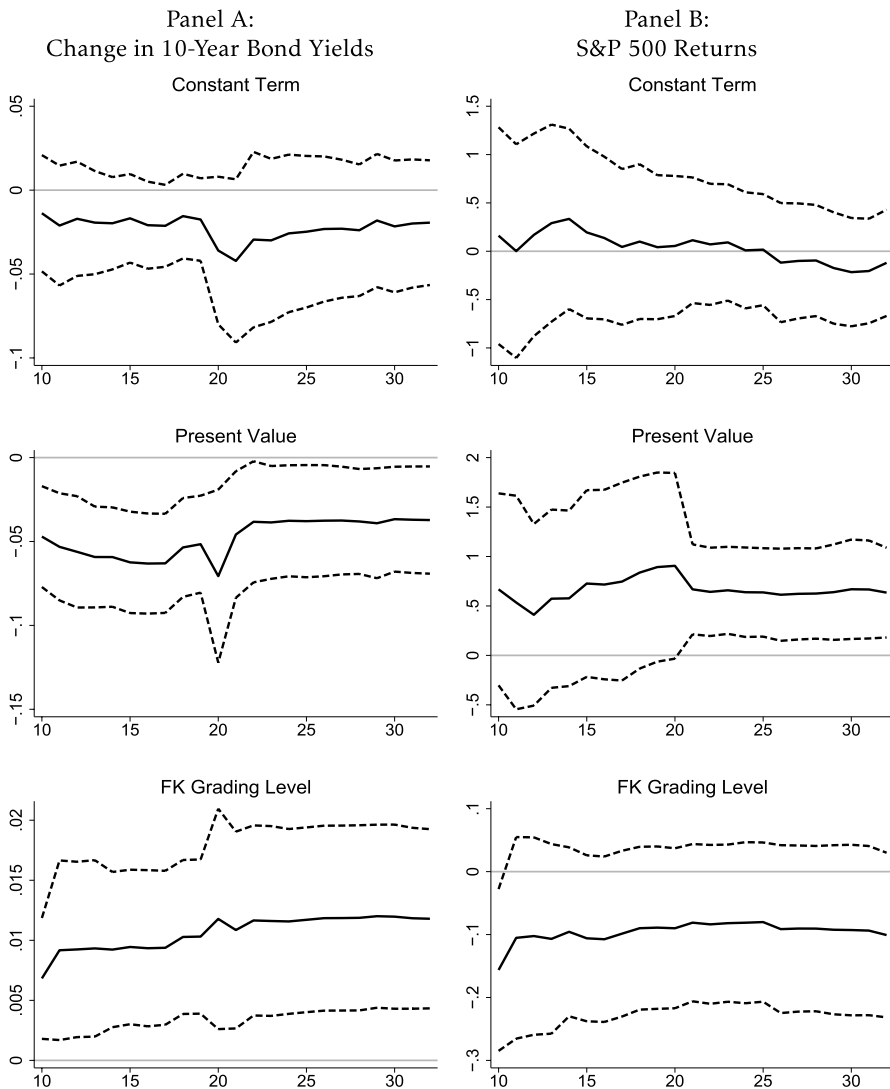
Fig. 5 Results of recursive estimations. *Notes:* Solid lines show the coefficients of a recursive estimation of Eq. (2) and the event window [**t−1**; **t+2**]. Dashed lines are the 95% confidence intervals based heteroskedasticity-consistent standard errors. y-axis: Size of the coefficients and the confidence intervals. x-axis: Last bill considered in the recursive estimation (see also Table 12)

Declarations

References

Abeler, J., & Jaeger, S. (2014). Complex tax incentives. *American Economic Journal: Economic Policy, 7*(3), 1–28.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2015). Democracy, redistribution, and inequality. *Handbook of Income Distribution, 2*, 1885–1966.

Acemoglu, D. (2006). A simple model of inefficient institutions. *Scandinavian Journal of Economics, 108*(4), 515–546.

Aidt, T. S. (2003). Economic analysis of corruption: a survey. *Economic Journal, 113*(491), F632–F652.

Ardagna, S. (2009). Financial markets' behavior around episodes of large changes in the fiscal stance. *European Economic Review, 53*(1), 37–55.

Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy, 4*(2), 1–27.

Baake, P., Borck, R., & Loffler, A. (2004). Complexity and progressivity in income tax design: Deductions for work-related expenses. *International Tax and Public Finance, 11*, 299–312.

Bacher, H. U., & Brülhart, M. (2013). Progressive taxes and firm births. *International Tax and Public Finance, 20*(1), 129–168.

Benoit, K., Munger, K., & Spirling, A. (2019). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science, 63*(2), 491–508.

Bloomfield, R. (2008). Discussion of 'annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2–3), 248–252.

Bradbury, M. E., Hsiao, P. K., & Scott, T. (2020). Summary annual reports: Length, readability and content. *Accounting and Finance, 60*(3), 2145–2165.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review, 99*(4), 1145–1177.

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

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies, 33*(4), 1673–1736.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance, 64*(2), 709–749.

Ehrmann, M., & Talmi, J. (2020). Starting from a blank page? Semantic similarity in central bank communication and market volatility. *Journal of Monetary Economics, 111*, 48–62.

Fazzari, S. M., Morley, J., & Panovska, I. (2015). State-dependent effects of fiscal policy. *Studies in Nonlinear Dynamics and Econometrics, 19*(3), 285–315.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233.

Gaertner, F. B., Hoopes, J. L., & Williams, B. M. (2020). Making only America great? Non-US market reactions to US tax reform. *Management Science, 66*(2), 687–697.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review, 70*(3), 393–408.

Hayo, B., Henseler, K., Rapp, M. S., & Zahner, J. (2022). Complexity of ECB communication and financial market trading. *Journal of International Money and Finance, 128*, 102709.

Hayo, B., & Mierzwa, S. (2020). "Stock market reactions to legislated tax changes: Evidence from the United States, Germany, and the United Kingdom." MAGKS Joint Discussion Paper Series in Economics No 47-2020.

Hirst, D. E., & Hopkins, P. E. (1998). Comprehensive income reporting and analysts' valuation judgments. *Journal of Accounting Research, 36*, 47–75.

Hopkins, P. E. (1996). The effect of financial statement classification of hybrid financial instruments on financial analysts' stock price judgments. *Journal of Accounting Research, 34*, 33–50.

IMF & OECD (2019). *Progress report on tax certainty. IMF/OECD Report for the G20 Finance Ministers and Central Bank Governors.*

Jansen, D. J. (2011). Mumbling with great incoherence: Was it really so difficult to understand Alan Greenspan? *Economics Letters, 113*(1), 70–72.

Jin, G. Z., Luca, M., & Martin, D. (2022). Complex disclosure. *Management Science, 68*(5), 3236–3261.

Kaplow, L. (1996). How tax complexity and enforcement affect the equity and efficiency of the income tax. *National Tax Journal, 49*(1), 135–150.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). "Derivation of new readability formulas for Navy enlisted personnel." *Naval Technical Training Command Millington TN Research Branch.*

Lee, Y. J. (2012). The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research, 29*(4), 1137–1170.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2–3), 221–247.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual Analysis, dictionaries, and 10-ks. *Journal of Finance, 66*(1), 35–65.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance, 69*(4), 1643–1671.

Lu, K. (2022). Overreaction to capital taxation in saving decisions. *Journal of Economic Dynamics and Control, 144*, 104541.

Lundholm, R. J. (1991). Public signals and the equilibrium allocation of private information. *Journal of Accounting Research, 29*(2), 322–349.

Mertens, K., & Ravn, M. O. (2010). Measuring the impact of fiscal policy in the face of anticipation: A structural VAR approach. *Economic Journal, 120*(544), 393–413.

Mertens, K., & Ravn, M. O. (2012). Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks. *American Economic Journal: Economic Policy, 4*(2), 145–181.

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

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *Accounting Review, 85*(6), 2107–2143.

Parker, J. A. (1999). The reaction of household consumption to predictable changes in social security taxes. *American Economic Review, 89*(4), 959–973.

Perotti, R. (2012). The effects of tax shocks on output: Not so large, but not small either. *American Economic Journal: Economic Policy, 4*(2), 214–237.

Poterba, J. M. (1988). Are consumers forward looking? Evidence from fiscal experiments. *American Economic Review, 78*(2), 413–418.

Romer, C. D., & Romer, D. H. (2009). "A narrative analysis of postwar tax changes," https://www.aeaweb.org/aer/data/june2010/20080421_app.zip.

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review, 100*(3), 763–801.

Rose-Ackerman, S. (Ed.). (2007). *International Handbook on the Economics of Corruption*. Cheltenham: Edward Elgar Publishing.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy, 2*(3), 180–212.

Schoonvelde, M., Brosius, A., Schumacher, G., & Bakker, B. N. (2019). Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political speeches. *PLOS One, 14*(2), e0208450.

Schuck, P. H. (1992). Legal complexity: Some causes, consequences, and cures. *Duke Law Journal, 42*, 1–52.

Souleles, N. S. (1999). The response of household consumption to income tax refunds. *American Economic Review, 89*(4), 947–958.

Souleles, N. S. (2002). Consumer response to the Reagan tax cuts. *Journal of Public Economics, 85*(1), 99–120.

Sims, C. (2005). "Rational inattention: A research agenda." Discussion Paper No. 34/2005, Deutsche Bundesbank.

Tan, H.-T., Wang, E. Y., & Zhou, B. (2014). When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research, 52*(1), 273–302.

Tullock, G., Brady, G. L., & Seldon, A. (2002). "*Government failure: A primer in public choice*." Washington. D.C.: Cato Institute.

US Comptroller General (1978). "*Further simplification of income tax forms is needed and possible*." GGD-78-74, Washington, D.C.: Government Printing Office, 3-4.

Wagner, A. F., Zeckhauser, R. J., & Ziegler, A. (2018). Company stock price reactions to the 2016 election shock: Trump, taxes, and trade. *Journal of Financial Economics, 130*(2), 428–451.

Warskett, G., Winer, S. L., & Hettich, W. (1998). The complexity of tax structure in competitive political systems. *International Tax and Public Finance, 5*, 123–151.

You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies, 14*(4), 559–586.