IMK
Macroeconomic Policy Institute

The IMK is an institute of the Hans-Böckler-Stiftung

WORKING PAPER

Nr. 169 · May 2016 · Hans-Böckler-Stiftung

# TOP-DOWN VS. BOTTOM-UP?

# RECONCILING THE EFFECTS OF TAX AND TRANSFER SHOCKS ON OUTPUT

**May 25, 2016**

Sebastian Gechert[1] Christoph Paetz[2] Paloma Villanueva[3]

## ABSTRACT

Using the bottom-up approach of Romer and Romer (2010), we construct a rich narrative dataset of net-revenue fiscal shocks for Germany by reconstructing and extending the tax shock series of Hayo and Uhl (2014) and coding a shock series for social security contributions, benefits and transfers. Based on quarterly data for 1974q1 to 2013q4 we estimate the multiplier effects of shocks to net-revenues, taxes, social security contributions and benefits in a proxy SVAR framework (Mertens and Ravn 2013) and compare them with estimates of the top-down identification inspired by Blanchard and Perotti (2002). We find multiplier effects of net-revenue components for Germany between 0 and 1 for both the top-down and bottom-up approaches. These estimates are on the lower end of the scale given in the literature and we discuss the differences.

[1] Sebastian Gechert, Corresponding author. Macroeconomic Policy Institute (IMK), Hans-Boeckler-Strasse 39, 40476 Düsseldorf, Germany. Email: sebastian-gechert@boeckler.de.

[2] Christoph Paetz, Macroeconomic Policy Institute (IMK).

[3] Paloma Villanueva, Banco de España. The views expressed in this paper do not necessarily reflect the views of the affiliated institutions.

We thank Michael Hachula, Katja Rietzler, Sven Schreiber, Johannes Steffen, Thomas Theobald, Achim Truger, Matthias Uhl and Rudolf Zwiener for helpful discussions. All remaining errors are ours.

# Top-Down vs. Bottom-Up? Reconciling the Effects of Tax and Transfer Shocks on Output

Sebastian Gechert* Christoph Paetz† Paloma Villanueva‡

May 25, 2016§

**Abstract.** Using the bottom-up approach of Romer and Romer (2010), we construct a rich narrative dataset of net-revenue fiscal shocks for Germany by reconstructing and extending the tax shock series of Hayo and Uhl (2014) and coding a shock series for social security contributions, benefits and transfers. Based on quarterly data for 1974q1 to 2013q4 we estimate the multiplier effects of shocks to net-revenues, taxes, social security contributions and benefits in a proxy SVAR framework (Mertens and Ravn 2013) and compare them with estimates of the top-down identification inspired by Blanchard and Perotti (2002). We find multiplier effects of net-revenue components for Germany between 0 and 1 for both the top-down and bottom-up approaches. These estimates are on the lower end of the scale given in the literature and we discuss the differences.

**Keywords.** Narrative Record Identification; Action-Based Approach; Fiscal Multipliers; Revenue Elasticities

**JEL classification.** E62, H20, H30

*Corresponding author. Macroeconomic Policy Institute (IMK), Hans-Boeckler-Strasse 39, 40476 Düsseldorf, Germany. Email: sebastian-gechert@boeckler.de. Tel: +49 211 7778 306.

†Macroeconomic Policy Institute (IMK)

‡Banco de España. The views expressed in this paper do not necessarily reflect the views of the affiliated institutions.

§We thank Michael Hachula, Katja Rietzler, Sven Schreiber, Johannes Steffen, Thomas Theobald, Achim Truger, Matthias Uhl and Rudolf Zwiener for helpful discussions. All remaining errors are ours.

## 1. Introduction

The size of fiscal multipliers, in particular for changes in public revenues, is still hotly debated. Since headline budgetary figures are prone to considerable endogeneity with respect to cyclical fluctuations, structural shifts of the tax base and one-off events, they do not lend themselves directly to policy analysis. When measuring the effects of discretionary fiscal policy changes on output, the discussion in the empirical literature of recent years thus centers around the underlying approach to identify those fiscal policy shocks that can be deemed exogenous.

Until 2010 the literature mainly relied on the top-down approach of Blanchard and Perotti (2002) (BP henceforth). This approach imposes restrictions from prior information on tax elasticities and implementation lags, in order to separate cyclical components of the budget from the exogenous discretionary ones in a structural VAR (SVAR). It allows identifying fiscal policy shocks from the reduced form model and then using impulse-response functions (IRF) to estimate the impact on economic performance. BP and some papers with applications of their method to other countries (Tenhofen et al. 2010; Perotti 2005; Baum and Koester 2011) find revenue multipliers fairly close to one or below.

Since the imposition of budget elasticities is prone to considerable uncertainty (Caldara and Kamps 2013), the narrative approach of Romer and Romer (2010) proposed a new bottom-up strategy. Based on legislative texts, presidential speeches and congressional reports, they identify the size, timing, and motivation of major legislated tax changes for the United States and construct a tax shock series from these narrative information. This procedure should reduce the endogeneity bias resulting from uncertain budget elasticities and endogenous discretionary policy responses to other shocks. Romer and Romer (2010) find large tax multipliers between two and three. Cloyne (2011) for the United Kingdom and Hayo and Uhl (2014) for Germany conduct bottom-up tax shock series as well and report similar results, regardless of country specifics.

These figures obviously do not square well with the benchmark BP identification. Favero and Giavazzi (2012) argue that the discrepancy between Romer and Romer (2010) and BP is not due to the different identification, but to the different estimation techniques. Romer and Romer (2010) used their tax shock series and estimated the effects on output in an ARDL model, while the results by BP stem from a VAR model. Employing the Romer and Romer (2010) shocks as an exogenous series in a VAR, and using dynamic multiplier functions, Favero and Giavazzi (2012) find multipliers in line with BP for the US. Hayo and Uhl (2014) as well as Cloyne (2011), however, do find tax multipliers between two and three with the Favero and Giavazzi (2012) specification, adding lags of the shock series. Perotti (2012) compares several specifications of the econometric model for US data, employing an IV approach to account for potential bias from other latent structural shocks, finding multiplier effects in the middle of the two extremes at about 1.5.

Mertens and Ravn (2013) provide another attempt to explain the differences for the US case. They feed the narrative identification process into the SVAR framework, by using the narrative shocks as proxies for the latent structural shocks of the VAR model and identifying the value of the impact multiplier directly via an IV estimation. This approach has the appeal that estimates of top-down and bottom-up multipliers are based on the same reduced-form transmission mechanism. Moreover, it allows for likely measurement error regarding the shock sizes. The tax multiplier found is in line with the one from Romer and Romer (2010) in the long run and already quite high on impact. Mertens and Ravn (2014) (MR henceforth) in a very thorough analysis point out that their higher multiplier results (as compared to Romer and Romer (2010)) can be attributed to a superior control for possible measurement error and fiscal foresight. Comparing their results to the much lower BP estimates, they argue that the latter are due to an underestimation of revenue elasticities, and adduce some indicative evidence in favor of their case.

In this paper we discover another possible explanation why multipliers from top-down and bottom-up approaches might differ. So far the bottom-up identification literature concentrated on estimating effects of tax shocks, while studies following the top-down BP method usually estimate net-revenue multipliers, that rest upon a complete measure of net-revenues, including taxes plus social security contributions minus benefits and transfers. Hence, the discrepancy in the results between the top-down and bottom-up identification could be due to the fact that the bottom-up construction of the shock series excludes changes in social security contributions, benefits and transfers that should be taken into account to provide a more complete picture of a fiscal shock series on the revenue side. Insofar as social security shocks and tax shocks imply different multipliers, this may drive the incompatible results. Romer and Romer (2014) for example find strong transfer multipliers on impact that diminish quickly, as opposed to their Romer and Romer (2010) tax multipliers that are low on impact and grow substantially within a 3 year horizon. Moreover, as far as tax changes and those for social security happen to have concurrent effects with interfering shock series, estimating tax multipliers without controlling for social security may yield biased results.

We provide two central innovations. First, exploiting official historical records of the German Bundestag and Bundesrat, the Federal Ministry of Labour and Social Affairs and the German statutory pension insurance scheme, we construct a series of legislated social security shocks for Germany. The dataset covers major changes in transfers and social security contributions for pensions, health care, long-term care and unemployment insurance on the German federal level for a quarterly time series spanning 1970 to 2013. We add this constructed narrative social security series to an updated version of the existing tax shock series of Hayo and Uhl (2014) for Germany thus providing a rich narrative record of net-revenues that also allows estimating multipliers for subcomponents of net-revenues.

Second, we feed the shock series into the proxy SVAR specification of MR and compare the structural impulse-response functions to those from a top-down SVAR estimation for Germany that uses the latest official figures of revenue elasticities as identifying restrictions. This enables a comparison of bottom-up and top-down multipliers based on the lowest possible degree of model friction. Moreover, we thoroughly discuss differences of our findings to the existing narrative estimates in the literature.

We find that using a rich narrative dataset for overall net-revenues within the MR specification yields multipliers of about 0.5 for Germany on impact and decreasing slowly within a time horizon of 5 years. This is much lower than other bottom-up estimates, but very close to our multiplier estimates from the top-down approach. We can thus reconcile the estimated revenue multipliers of the top-down and bottom-up approaches with multipliers on the lower end of the scale given in the literature. As a mirror image, the implied revenue elasticity to changes in GDP resulting from the bottom-up identification is fairly close to the one imposed for the top-down approach. Hence, the finding of the existing literature that bottom-up identified revenue multipliers and elasticities are much larger than conventional top-down estimates does not seem to be generalizable.

As opposed to the hypotheses laid out above, these findings are not driven by the use of different revenue categories or the omission of correlated shocks: First, social security shocks and tax shocks are largely uncorrelated in our sample, such that their impact on GDP can be estimated separately without serious bias. Second, when analyzing the revenue components separately, results for top-down and bottom-up estimations are fairly close to each other and close to those from the compound net-revenue series within a range of zero to one along the 5-year horizon. Tax multipliers are below 0.5 for our baseline specification and do not exceed 1.2 in any of the alternative specifications. Shocks to social security revenues imply multipliers that are close to one on impact and die out quickly. Changes to social security expenditures yield multipliers below one, which are somewhat more persistent within the five-year horizon. As a general conclusion, we find that expansionary tax and social security changes have a positive but only limited short-to-medium-term impact on GDP for Germany.

The much larger tax multipliers of Hayo and Uhl (2014) for Germany seem to result from their different econometric approach: in contrast to the proxy SVAR, it (i) does not allow for uncertainty concerning the size of the narrative revenue shocks and (ii) requires orthogonality of all included lags of the shock series with other latent structural shocks, thus being more restrictive than the MR approach (Mertens and Ravn 2013). Our much lower multipliers as compared to the US proxy SVAR estimation of MR are to a considerable extent driven by an alternative choice regarding the scaling of shocks, which about halves the results for the US sample, but only insignificantly affects our results for Germany. An additional channel explaining the differences may be the stronger imports-to-GDP ratio of the German economy as compared to the US. Fiscal foresight does not seem to drive the differences as our results remain largely unaffected.

The remainder of the paper is organized as follows. In Section 2 we describe the construction of the narrative shock series and examine them. Section 3 presents the econometric framework and the opposing approaches to identification. Afterwards, we present our findings regarding the multiplier effects of both the bottom-up and top-down estimation in Section 4 and discuss them in relation to findings of the existing literature. We test their robustness in Section 5. The final section concludes.

## 2. Constructing and Examining the Shock Series

This section lays out how we identified the exogenous shock series for net-revenue changes following the bottom-up approach. For any judgment calls, we closely stick to Hayo and Uhl (2014), in order to reach the highest possible degree of comparability. A detailed description of the construction of our social security shock series can be found in the companion paper (Gechert et al. 2016), which also complements the tax shock narrative of (Uhl 2013).

In contrast to the construction of the tax shock series, expected impacts of discretionary policy changes in benefits, transfers and social security contributions are not listed in the annual budgetary report of the Federal Ministry of Finance (*Bundesfinanzberichte*). In order to identify major changes to social security and transfer legislation, we therefore rely on chronicles from the Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales 2011) and various *Sozialberichte*, the chronicle of the German statutory pension insurance scheme (German Statutory Pension Insurance 2011: 267-308) as well as Steffen (2013), who provides a chronicle of major legislations for all subdivisions of social security. From these chronicles, we set up a list of major legislations for pensions, health care, long-term care and unemployment insurance at the German federal level for the period 1970 to 2013. For each law listed in the chronicles, we then filed through draft legislations, bills, parliamentary protocols and speeches in order to collect information regarding (i) the underlying motivation, (ii) the dates of the legislative process and (iii) the prospective financial impact.

(i) A central advantage of the bottom-up approach is that one can readily select discretionary measures and separate them from all automatic fluctuations of the budget. However, discretionary measures can still be endogenous reactions to changing circumstances, which would invalidate the causal interpretation of estimation results. Following Romer and Romer (2010) we assign to each law an exogenous or endogenous underlying motivation. In line with Hayo and Uhl (2014), we classify those measures as endogenous, which are either driven by policies that contemporaneously affect other budgetary positions with interfering effects but outside the information set of the narrative (spending-driven or revenue-driven motivation), countercyclical policies or reactions to other macroeconomic shocks (like financial crises, oil price shocks, etc.). Refraining to consider these measures in the shock series should rule out likely biases from omitted variables. The relevant exogenous changes that lend themselves to a causal interpretation with respect to short-run multiplier effects are those that are motivated by attempts to long-term budgetary consolidation, structural or ideological reasons or rulings of the court.[1]

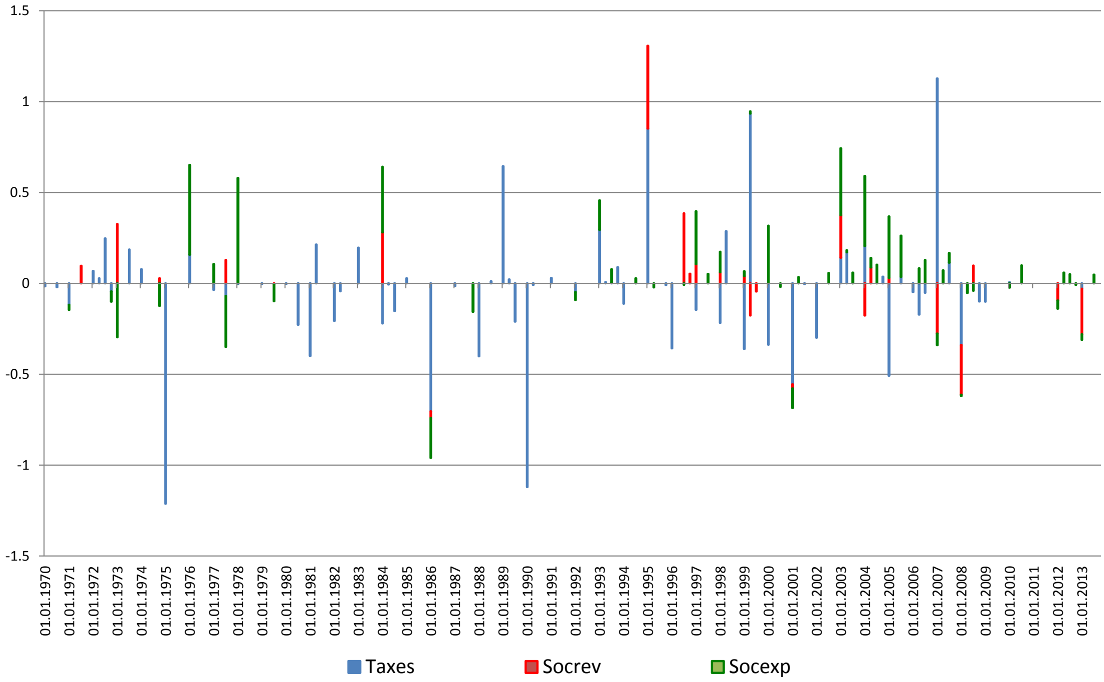
(ii) From the information of the law, we are able to detect the timing of the implementation of a measure in order to determine the quarter of the shock in our data set. Similarly to Hayo and Uhl (2014), we take record of different implementation dates of individual measures within a law code if applicable and check whether they are temporary or permanent. In the event that the measure is of a temporary nature, the date of its expiration is recorded as well and provides the timing of the respective counter-shock (of the same size). When temporary measures are prolonged, a new shock with the new expiration date is included. Additionally, we also collect the announcement date of the legislation, which is uniform for all single measures of a law. The announcement date usually coincides with the publication of the law's first draft, which usually contains all relevant information concerning motivation, expected sizes of shocks and implementation dates.

(iii) The size of the shock and the economic relevance of each law is determined by its total expected full-year impact divided by annual nominal GDP in the year of the shock. The best available information on this impact is given in the drafts of each law. As a general caveat, it should be noted that these figures are ex-ante evaluations that are prone to uncertainty and probably to political bias.

We include all laws in the shock series with an expected total impact after full implementation above or just slightly below 0.1% of annual nominal GDP at the quarter the law was implemented. Furthermore, similar to Hayo and Uhl (2014) we include laws where substantial budgetary impacts of single measures are canceled out by each other or by temporary measures. Moreover, if a law with small changes is introduced contemporaneously with other substantial changes, we include its effect as well in order not to bias the impact of the substantial change. The size of shocks represents the

[1] For an extensive explanation of these categories see Romer and Romer (2010); Hayo and Uhl (2014).

Figure 1: Exogenous Shocks to Taxes and Social Security at Implementation Date (% GDP) (A positive sign indicates a consolidation shock)

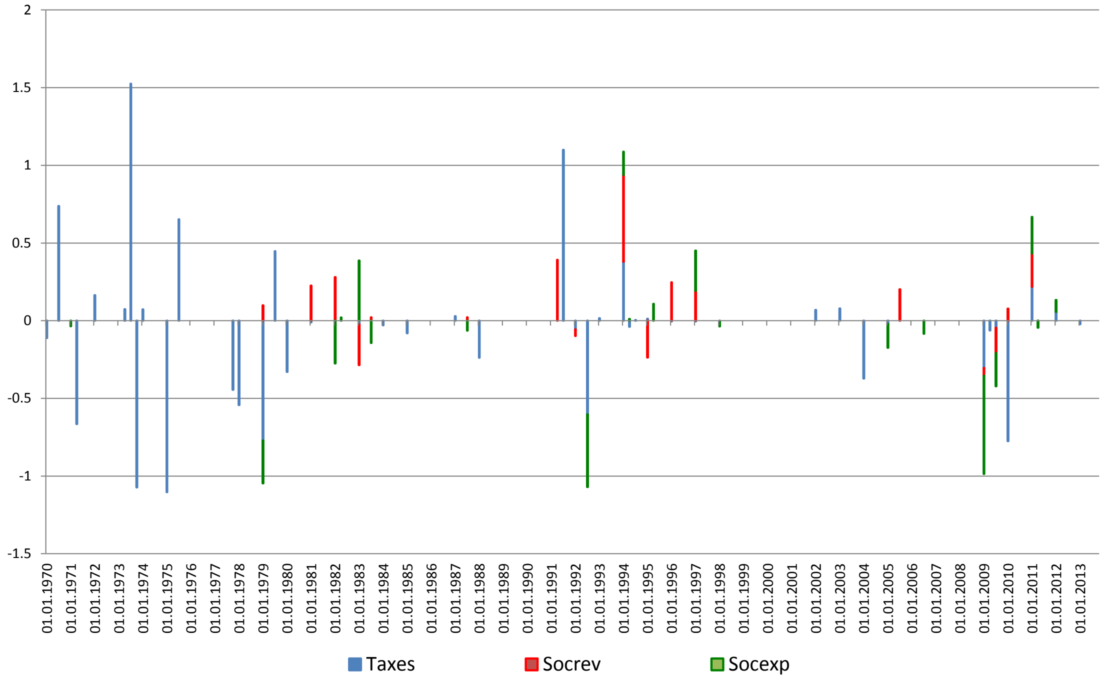



prospective annual financial impact after full implementation as a percentage of annual GDP, assuming no change in the tax base.

Figure 1 entails our constructed exogenous shock series at implementation dates. Tax shocks are reconstructed from the information in Hayo and Uhl (2014) and extended up to 2013q4. Without access to the original dataset, we drew information from the companion paper (Uhl 2013) and their original source, the *Bundesfinanzberichte*. The figures of the shock series look quite similar and we are able to closely reproduce their results (see Appendix B).

Apparently, there are some contemporaneous shocks for social security contributions (Socrev), benefits (Socexp) and taxes (Taxes), giving rise to the hypothesis that there could be interfering effects. Contemporary and lagged correlation (± 4 quarters), however, is quite low among the three series ($|cor(\tau_i, \tau_j)| < 0.1$). We therefore do not expect biased estimates of multipliers when using the net-revenue components separately. In

Figure 2: Endogenous Shocks to Taxes and Social Security at Implementation Date (% GDP) (A positive sign indicates a consolidation shock)




line with results by Romer and Romer (2010) and Cloyne (2013) the mean for taxes is slightly negative with -0.016 % of GDP (standard deviation: 0.22). The degree of volatility is similar to Romer and Romer (2010) and Cloyne (2013). The mean for Socrev is positive but very low with 0.006 (sd: 0.07) and for Socexp 0.018 (sd: 0.10), slightly positive as well. Volatility is lower than for taxes for both social security shock series. The mean of our full exogenous shock series for net-revenues is very low with 0.008 (sd: 0.26).

Figure 2 shows those shocks for Taxes, Socrev and Socexp which are endogenously motivated. The total endogenous series of net-revenues has a slightly negative mean of -0.031 % of GDP (sd: 0.51). Endogenous reactions to the economic development by tax policies were actively used in the 1970s and became a less important tool since the beginning of 1980s. Endogenous changes to the social security system were concentrated at the beginning of the 1980s, the mid 1990s and in response to the financial crises.

Table 1: Predictability of the shock series – Granger causality tests (based on lags 1 through 4 of growth rates of GDP, government expenditures and the respective net-revenue component.)

| | | Taxes | Socrev | Socexp | Netrev |
|---|---|---|---|---|---|
| Exo | $\chi^2$ | 4.601 | 12.479 | 11.601 | 9.358 |
| | $p(\chi^2)$ | 0.970 | 0.408 | 0.478 | 0.672 |
| Endo | $\chi^2$ | 26.459 | 21.514 | 17.934 | 18.243 |
| | $p(\chi^2)$ | 0.009 | 0.043 | 0.118 | 0.109 |

A major concern regarding the assumption of exogeneity of the shock series is fiscal foresight which may result in different information sets of agents and the econometrician, thereby ignoring possible reactions to predictable shocks that happen prior to the implementation of the law (Mertens and Ravn 2010; Ramey 2011). Table 1 captures the predictability of our shock series based on Granger causality tests against the lagged values of the macroeconomic series that we include in our estimation in Section 4 (GDP, government expenditures and revenues). We cannot reject that the exogenous shock series are not predictable from the included macroeconomic series. In contrast, the shocks classified as endogenous, seem to be predictable, even though the test statistic is borderline non-significant at the 10 percent threshold in the case of endogenous social security expenditures and overall net-revenues.

Dealing with the issue of fiscal foresight in more detail, we also present results based on shocks dated at the announcement date of the respective legislations in Section 5. Moreover, in line with MR, we also discuss the case of legislations, whose implementation follows swift after their announcement such that anticipatory effects can be largely ruled out.

## 3. Model, Data and Identification

To foster a rigorous comparison of the top-down and bottom-up identification in terms of revenue multipliers, the architecture of our VAR model is equal for both approaches:

$$\boldsymbol{\Gamma}(L)X_t = v + u_t \tag{1}$$

$$\boldsymbol{A}\boldsymbol{\Gamma}(L)X_t = \boldsymbol{A}v + \boldsymbol{B}\varepsilon_t \tag{2}$$

$$X_t = \begin{bmatrix} g_t & y_t & \tau_t \end{bmatrix}' \tag{3}$$

Equation (1) represents the reduced-form model, while (2) follows the structural representation of the AB-model in Lütkepohl (2006: 364). $\Gamma(L)X_t$ is a 4th-order lag polynomial of the $K$ (lagged) endogenous variables $X_t$ and their coefficients $\Gamma$. For our baseline estimation, all variables are in log-levels.

In line with BP and MR, $X_t$ includes the log of real per capita government spending on consumption and capital formation ($g_t$), the log of real per capita GDP ($y_t$) and the log of real per capita net-revenues ($\tau_t$) (taxes plus social security contributions minus transfers or the single components interchangeably). In an extended specification that is closer to Hayo and Uhl (2014), we add the log of the GDP deflator ($p_t$) and a nominal short term interest rate ($i_t$). Moreover, we estimate a specification in first differences.

Data for GDP and the GDP deflator are taken from the OECD Quarterly National Accounts and transformed to annualized levels. Levels prior to unification are extrapolated by means of West German growth rates. The budgetary data stem from the financial statistics of the Bundesbank and are cash-based ("Finanzstatistik"). Data for population are taken from the German Federal Statistical Office. All series are seasonally adjusted using X-12-Arima and the price adjustment is based on the GDP Deflator. The short-term nominal interest rate is obtained from the OECD Monthly monetary and financial statistics (MEI).

The effective sample spans 1974q1 to 2013q4, despite the availability of narrative information back to 1970q1, since fiscal quarterly series are only available from the beginning of 1974. Our narrative dataset that is described in Section 2 allows extending the sample period of Hayo and Uhl (2014), whose estimation is based on 1974q1 to 2010q2.

$v$ contains a constant, a linear time trend, a re-unification step dummy (1991q1-2013q4) and a financial crisis dummy (2009q1). $u_t$ is the $K \times 1$ vector of reduced-form disturbances, while $\varepsilon_t$ contains the $K \times 1$ structural-form shocks that are to be identified by either the top-down or bottom-up method. $\boldsymbol{A}$ and $\boldsymbol{B}$ are the $K \times K$ factorization matrices that contain the contemporaneous dependencies among the endogenous variables and the structural shocks, respectively. The relation between $u_t$ and $\varepsilon_t$ boils down to

$$u_t = \boldsymbol{A}^{-1}\boldsymbol{B}\varepsilon_t. \tag{4}$$

Solving this system of equations requires estimating the variance-covariance matrix $\Sigma_u$ of the reduced-form residuals. Without loss of generality, we assume ortho-normality of the structural shocks ($\varepsilon_t \sim (0, \Sigma_\varepsilon = I_K)$) and exploit the relation

$$\Sigma_u = \boldsymbol{A}^{-1}\boldsymbol{B}\Sigma_\varepsilon\boldsymbol{B}'(\boldsymbol{A}^{-1})' = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{B}'(\boldsymbol{A}^{-1})'. \tag{5}$$

Identification can be achieved by imposing $(K^2 + K(K-1)/2)$ restrictions on $\boldsymbol{A}$ and $\boldsymbol{B}$.

Following the BP approach, we first set the following technical zero and one restrictions:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\bar{\alpha}_{gy} & -\bar{\alpha}_{g\tau} \\ -\alpha_{yg} & 1 & -\alpha_{y\tau} \\ -\bar{\alpha}_{\tau g} & -\bar{\alpha}_{\tau y} & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} \beta_{gg} & 0 & \bar{\beta}_{g\tau} \\ 0 & \beta_{yy} & 0 \\ \beta_{\tau g} & 0 & \beta_{\tau\tau} \end{bmatrix} \tag{6}$$

The BP approach uses additional prior assumptions on budget elasticities and institutional settings for identification, where $\bar{(\cdot)}$ denotes a restricted parameter: (i) Leaving $\beta_{\tau g}$ unrestricted and setting $\beta_{g\tau} = 0$ implies that in the process of setting up the public budget, spending decisions are taken prior to revenue decisions, an assumption which has been shown to be robust for US data by BP. We also show the robustness of this choice for our sample in Section 5. (ii) Government direct spending (excluding transfers and interest) is assumed to be inelastic to GDP and taxes within a quarter ($\alpha_{gy} = \alpha_{g\tau} = 0$) and also tax revenues are assumed not to be driven by government spending over and above what has been said under (i), thus imposing $\alpha_{\tau g} = 0$.

(iii) The crucial assumption for estimating revenue multipliers with the top-down approach concerns the elasticity of revenues to GDP $\alpha_{\tau y}$. We determine $\alpha_{\tau y}$ for our different revenue categories based on the latest OECD estimates (Price et al. 2014). A detailed description can be found in Appendix A. The respective elasticities used for different revenue categories are given in row (1) of Table 2. According to these figures, the German tax system is slightly progressive. Social security adds to progressivity, such that the overall net-revenue budget strongly reacts to a change in GDP.

Caldara and Kamps (2012) show that within a reasonable range of $\alpha_{\tau y}$, not even the sign of the resulting multiplier can be robustly estimated, such that both negative and large positive multipliers can occur. The very nature of the BP approach for estimating revenue multipliers, however, rests upon the assumption of a certain value of $\alpha_{\tau y}$ that is imposed as a scalar without taking into account likely uncertainty around this figure. We test the sensitivity of our results to a range of values of $\alpha_{\tau y}$ in Section 5. Comparing the values with the estimates from the MR approach provides another useful test as to whether the restrictions are valid (see below).

Imposing the restrictions (i-iii) is sufficient for a just-identified model. Setting the $\alpha_{\tau y}$ value has the advantage that the contemporaneous reaction of GDP to changes in revenues $\alpha_{y\tau}$ can be left unrestricted and be determined by the data. Rows (4) and (5) of Table 2 list the implied elasticities of $y$ to a change in $\tau$ for our baseline estimation of

Table 2: Elasticities imposed and estimated for the BP and MR models in levels (L) and growth rates (G)

| | | Taxes | Socrev | Socexp | Netrev |
|---|---|---|---|---|---|
| $\alpha_{\tau y}$ | | | | | |
| (1) | BP imposed | 1.08 | 0.60 | -0.50 | 2.71 |
| (2) | MR implied L | 0.8 (0.48,1.11) | 0.75 (0.45,1.06) | -0.6 (-1,-0.21) | 2.2 (1.4,3.01) |
| (3) | MR implied G | 1.37 (0.99,1.75) | 0.56 (0,23,0.9) | -0.33 (-0.75,0.1) | 3.18 (2.22,4.14) |
| $\alpha_{y\tau}$ | | | | | |
| (4) | BP implied L | -0.12 (-0.22,-0.03) | -0.09 (-0.18,0) | 0.15 (0.07,0.23) | -0.11 (-0.16,-0.05) |
| (5) | BP implied G | -0.11 (-0.2,-0.02) | -0.09 (-0.18,0) | 0.11 (0.03,0.18) | -0.09 (-0.14,-0.04) |
| (6) | MR imposed L | -0.04 | -0.14 | 0.16 | -0.06 |
| (7) | MR imposed G | -0.19 | -0.08 | 0.07 | -0.10 |

95% confidence bounds for implied elasticities in parentheses.

the SVAR both in levels (L) and growth rates (G). These can be transformed into the impact revenue multipliers by re-scaling with the sample-average ratio of $\tau/y$ (in linear levels).

Turning to the bottom-up approach, identification is achieved by recording exogenous changes to tax and social security legislation, determining their timing, ex-ante impact on revenues and motivation, as described in Section 2. The crucial assumption is, that the conducted narrative shock series $m_t$ is orthogonal to other structural shocks, which basically would allow a direct dynamic regression of GDP on the shock series, like Romer and Romer (2010) did by using an ARDL model. In order to account for other feedback effects, the literature that followed employed a standard VAR of budgetary components, GDP and other macro variables, including (lags of) the narrative shock series as exogenous variables (Favero and Giavazzi 2012; Cloyne 2013; Hayo and Uhl 2014).

$$\mathbf{\Gamma}(L)X_t = v + \boldsymbol{\lambda}(L)m_t + w_t \tag{7}$$

They then proceed by estimating dynamic multiplier functions of GDP to a shock in the narrative series. However, these dynamic multiplier functions are not necessarily identical to the impulse-response functions (IRF) from a structural VAR. First, adding the shock series (and its lags) as exogenous regressor(s) implies a different reduced form VAR model than in equation (1). Second, using the narrative shocks as a direct replacement of the latent structural revenue shocks may be invalidated because of measurement error and judgment calls when setting up the narrative record. This makes an instrumented approach more appealing, as the latter requires only imperfect correlation between the narrative shock series and the latent structural shocks ($E[m_t \varepsilon_t^{\tau}] \neq 0$). Third, exogeneity requires the included lags of the shock series to be uncorrelated with other latent structural shocks collected in $w_t$.

We therefore follow the *proxy SVAR* approach of MR that takes account of these issues. MR use the same reduced form VAR model as in the BP approach.[2] Identification includes a three-step procedure: (i) The VAR is estimated in reduced form without the shock series. (ii) The residuals $u_t^i,\ i \in g, y$ are regressed on $u_t^{\tau}$ using the shock series $m_t$ as the instrument.

$$\hat{u}_t^i = \mu^i + \alpha_{i\tau}^{IV} \bar{u}_t^{\tau} + \zeta_t^i \tag{8}$$

$$\hat{u}_t^{\tau} = \mu^{\tau} + \gamma m_t + \zeta_t^{\tau} == \mu^{\tau} + \bar{u}^{\tau} + \zeta_t^{\tau} \tag{9}$$

(iii) The coefficients $\alpha_{i\tau}$ are then imposed on the $\boldsymbol{A}$ matrix (with $\alpha_{\tau\tau} = 1$ by definition), if necessary, alongside with other identifying restrictions. The factorization matrices read

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\bar{\alpha}_{gy} & -\bar{\alpha}_{g\tau} \\ -\alpha_{yg} & 1 & -\bar{\alpha}_{y\tau} \\ -\bar{\alpha}_{\tau g} & -\alpha_{\tau y} & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} \beta_{gg} & 0 & \bar{\beta}_{g\tau} \\ 0 & \beta_{yy} & 0 \\ \beta_{\tau g} & 0 & \beta_{\tau\tau} \end{bmatrix} \tag{10}$$

Again we restrict $\alpha_{gy} = \alpha_{\tau g} = \beta_{g\tau} = 0$. The crucial difference to BP is that $\alpha_{i\tau}$ are determined by the IV regression, while leaving the critical revenue elasticity $\alpha_{\tau y}$ unrestricted. For comparison, the imposed and implied elasticities for the MR approach can

[2]Note that MR actually employ the $\mathfrak{B}$ model of factorization, as will be discussed below.

Table 3: Relevance and reliability of the instrument

| | | Taxes | Socrev | Socexp | Netrev |
|---|---|---|---|---|---|
| L | $F(u_t^\tau, m_t)$ | 11.025 | 3.091 | 6.270 | 3.689 |
| | $p(F)$ | 0.001 | 0.081 | 0.013 | 0.057 |
| | $R^2(\varepsilon_t^\tau, m_t)$ | 0.170 | 0.150 | 0.123 | 0.054 |
| G | $F(u_t^\tau, m_t)$ | 7.456 | 3.977 | 6.084 | 4.324 |
| | $p(F)$ | 0.007 | 0.048 | 0.015 | 0.039 |
| | $R^2(\varepsilon_t^\tau, m_t)$ | 0.185 | 0.170 | 0.089 | 0.072 |

be found in rows (2), (3), (6) and (7) of Table 2. These estimates stem from the SVAR estimations in levels and growth rates, respectively. The figures are particularly close for social security revenues and expenditures and borderline insignificantly different at 95% confidence bounds for tax revenues and the compound net-revenues. Thus, as opposed to MR, for the German case we cannot reject that the top-down and bottom-up identification lead to equivalent multipliers and elasticities for various revenue components. The similarity of the IRFs presented in Section 4 mirrors this finding.

Table 3 tests the relevance and reliability of the instrument in the 2SLS-regression (8) both for the specification of the SVAR in levels (L) and growth rates (G). F-tests and respective p-values for the first stage show the relevance of the instrument. In line with MR, reliability of the narrative instrument $m_t$ for the true underlying revenue shock series $\varepsilon_t^\tau$ is derived by regressing the estimated structural shocks $\hat{\varepsilon}_t^\tau$ on the non-zero observations of $m_t$, which should asymptotically be equivalent to the reliability of the instrument (Mertens and Ravn 2013). For tax revenues, the instrument seems relevant and $m_t$ shocks have some predictive power for $\hat{\varepsilon}_t^\tau$. The test results are somewhat weaker for the other revenue components, thus their multiplier results should be interpreted with more caution.

For the evaluation of multiplier effects, one first has to transform the usual 1-SD shocks to 1% of GDP changes, which is usually done by normalizing with the sample-average ratio of taxes to GDP. Second and more critically, one has to take a stance on the definition of the revenue shock, namely, as to whether it accounts for contemporaneous feedback via automatic stabilizers: Either it can be interpreted (a) as an increase in projected tax liabilities of 1% of GDP ($\varepsilon^{\tau} = 1$), excluding the feedback on the tax base or (b) as an increase of effective revenues of 1% of GDP including the feedback ($u^{\tau} = 1$). MR follow option (b) by employing the $\mathfrak{B}$ model of factorization (where $\mathfrak{B} = A^{-1}B$), and – via the IV estimation – identifying the $\mathfrak{B}_{\tau}$ column vector up to a scaling factor. They deliberately choose the scaling factor such that it implies a structural shock size ($\varepsilon^{\tau}$) that corresponds to $u^{\tau} = 1$, and is transformed into 1% of GDP in collected revenues after feedback.

Using the AB model instead enables us to account for both options (a) and (b). By the IV estimation, we determine the $\mathbf{A}_{\tau}$ column vector. At the expense of setting an additional identifying restriction[3] we then estimate $\beta_{\tau\tau}$ by which we scale the IRFs to a shock of $\varepsilon^{\tau}$ equal to 1% of GDP, thus excluding any initial feedback via $u^{y}$ or $u^{g}$.

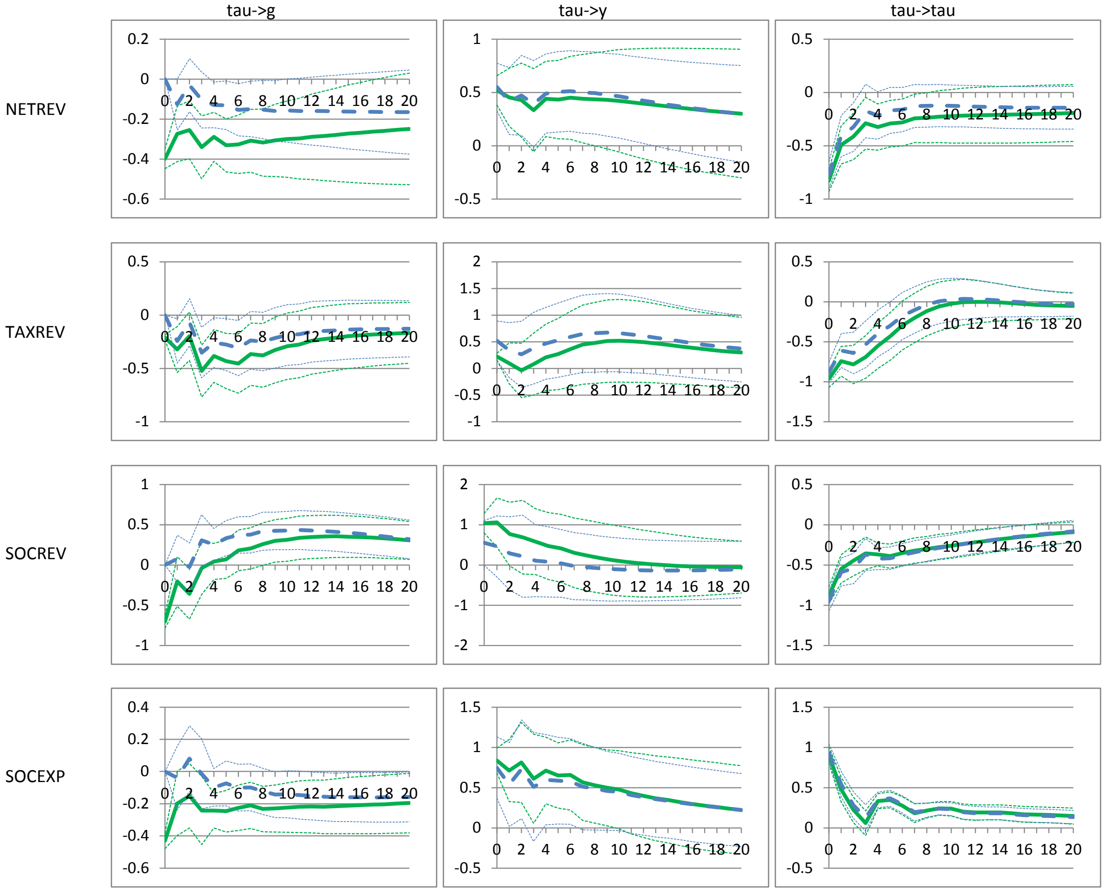
With reasonable signs for $\alpha_{\tau y}(> 0)$ and $\alpha_{y\tau}(< 0)$, the reported GDP reaction is stronger for (b) than for (a), as it requires an increase in tax liabilities of more than 1% of GDP to raise 1% of GDP in effective revenues. Thus, the shock size $\varepsilon^{\tau}$ is inflated to arrive at ($u^{\tau} = 1$). Note that the difference can be huge when the absolute values of $\alpha_{\tau y}$ and $\alpha_{y\tau}$ are big. Since the strength of the feedback is endogenous to the results, we prefer to compare the pure multiplier effects of ($\varepsilon^{\tau} = 1$) and follow option (a) for our baseline estimates, but also test the alternative choice (b).

## 4. Results

We now estimate the responses of the endogenous variables to an expansionary shock ($-\varepsilon_t^{\tau}$) to the respective revenue series, that is, either a relief in net revenues (NETREV), taxes (TAXES) or social security contributions (SOCREV), or an increase in benefits and transfers (SOCEXP). Shocks are sized to 1% of GDP of prospective revenues (or

[3] Our preferred choice is $\alpha_{gy} = 0$, in line with our top-down specification.

Figure 3: Impulse-responses for BP (dashed blue) and MR (solid green) identification after expansionary shock of 1% of GDP to various net-revenue components (tau), log levels, 2-SD error bands




expenditures) without feedback on the respective tax base for a horizon of 20 quarters. The error bands are 2-SD analytic confidence intervals. Figure 3 shows the IRFs of our baseline specification in levels.

Using both the BP top-down and MR bottom-up methodology and identification, we find plausible multiplier effects on $y$ between zero and one for our different net-revenue components. Generally, the dynamics of the IRFs are akin for the BP and MR approaches, which is not surprising, given the identical reduced form model in use. Only the impact values differ somewhat in line with Table 2. As opposed to MR and Hayo and Uhl (2014) we do not find strong tax multipliers for Germany with the bottom-up identification. Tax multipliers are even moderately lower on impact for the MR case than for the BP case and are not significantly different from zero after the impact quarter. Multipliers for social security components are to some extent higher. The GDP response to changes in social security contributions fades relatively quickly, while social security expenditures imply somewhat more persistent effects. The compound net-revenue shocks on average lead to multipliers of around 0.5 with only slightly different effects for the BP and MR approaches.

What accounts for the difference between the Hayo and Uhl (2014) estimates of huge tax multipliers of around 2.5 and our rather small multipliers? It does not seem to be driven by the shock series itself: even though our sample is extended to 2013q4 and we did not have access to the precise dataset, we are able to reproduce their findings when applying the same VAR specification as laid out in the Appendix B. However, two methodical distinctions stand out: First, Hayo and Uhl (2014) find their large multipliers for a specification in first differences. When they estimate their VAR in log-levels, the peak multiplier of tax changes shrinks to about 1.6 (and to about 1.9 in our replication of their level estimate). Nevertheless, since this compares to our tax multiplier of 0.5 at the peak, the difference is still economically and statistically significant. Similarly, our estimate in first differences implies a peak point estimate of the GDP response of 1.2, compared to the 2.4 we find for the Hayo and Uhl (2014) baseline specification. Second, the specification of Hayo and Uhl (2014) requires perfect correlation between the narrative shock series and the latent structural shocks and exogeneity of the included lags of the shock series. Ruling out other possible differences, the high tax multipliers of Hayo and Uhl (2014) seem to be driven by these stricter assumptions, which may not be valid.

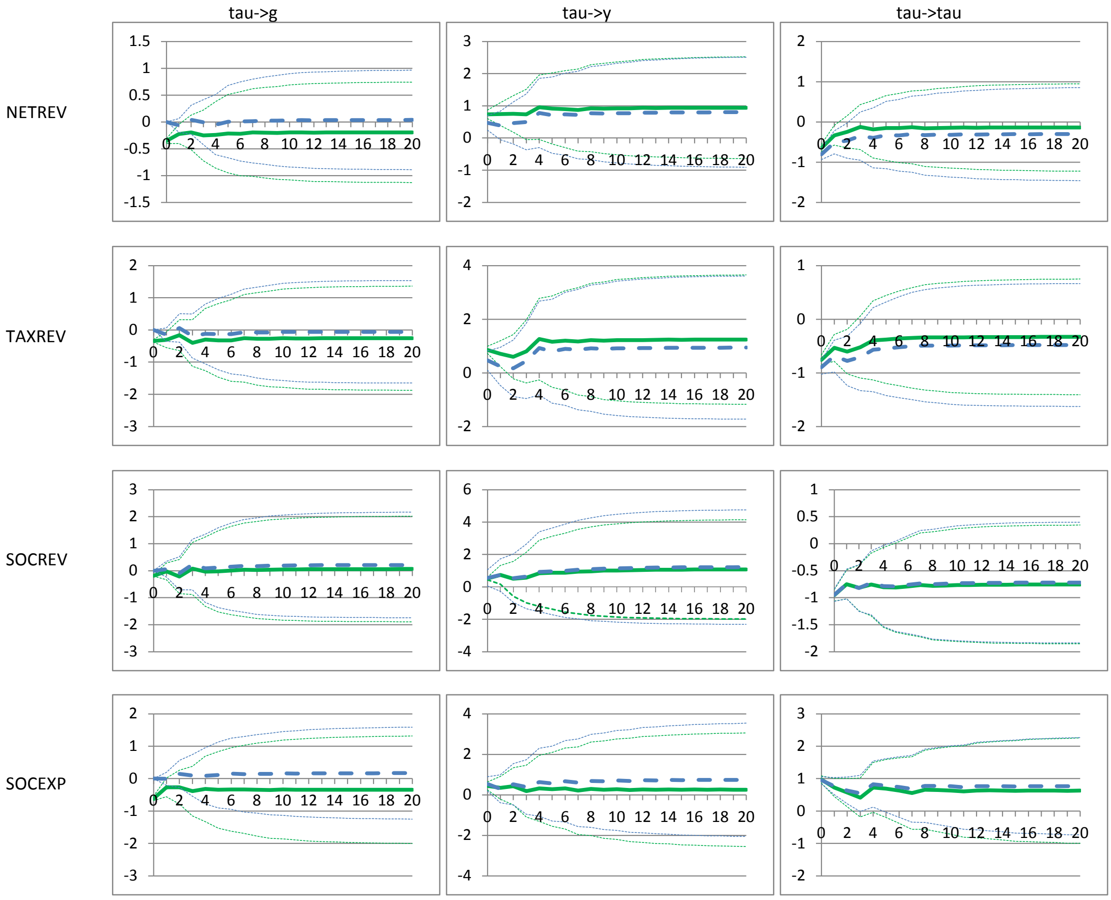
What accounts for the much lower multipliers in our case as compared to Mertens and Ravn (2014)? The modelling framework is fully consistent and we are able to reproduce their results for their US sample with our slightly different set of identifying restrictions (see Figure 7 in Appendix B). Three differences are apparent:

(i) MR base their estimations on non-anticipated tax shocks of the Romer and Romer (2010) US series, by including only those observations, where announcement and implementation date of the law are less than 90 days apart. For our German dataset, such fast legislations are rare. Based on the few non-zero observations left, the relevance of the instrument regression becomes rather weak and the first stage regression shows an insignificant correlation between the net-revenue time series and the respective shock series. We thus tested a shock series for net-revenues with a 180-days threshold. The response of GDP (Figure 5a, blue dashed line MRnant) is somewhat lower than in our baseline case that includes all exogenously motivated shocks, but the difference becomes insignificant soon after the first quarters. This finding is reasonable, given that in the case of an anticipated tax relief business activity is likely to be postponed until after its implementation. As opposed to MR we did not use the non-anticipated shocks as our baseline case, because the low number of shocks left would prevent a reasonable investigation of the effects of the separate net revenue components. In order to test the robustness of this choice we also perform another test of fiscal foresight proposed by Cloyne (2013). The green solid line (MRanno) of the GDP response is based on a shock series dated at the announcement of the legislations. The result is remarkably close to the one for non-anticipated effects. Again, there is a lower response at the announcement date as compared to the baseline measure at implementation date, which is plausible if agents postpone activity up until the tax relief is implemented. As compared to the US case covered in MR, where non-anticipated tax shocks imply much higher multipliers than using the full set of the Romer and Romer (2010) tax shocks, fiscal foresight does not seem to be as relevant in our case. After all, controlling for anticipation even increases the difference between MR's and our results.

(ii) MR identify the structural shocks up to a scaling factor, and deliberately scale their shocks such that they mimic an effective change in tax revenues of 1% of GDP after macroeconomic feedback effects; our approach allows computing the strength of the feedback at the expense of an additional identifying restriction and thus separating the initial change in tax rates, the subsequent multiplier effect with its impact on the tax base, and the latter's feedback on actual tax revenues (see options (a) and (b) in Section 3 again). Due to the comparably big values for $\alpha_{\tau y}$ and $\alpha_{y\tau}$ in the MR data, the feedback is quite strong. When we estimate the GDP effect of a shock to the narrative series of size 1% of GDP for the US sample used in MR (see AppendixB), we find a pure tax multiplier of about 1 on impact and 1.6 at the peak. This pure multiplier effect is only about half as strong as the ex post multiplier reported by MR. The pure multiplier feeds back on actual tax revenues which only decrease by about 0.5 in the initial quarter after the shock. In other words, the strong multiplier and tax elasticity of MR imply a self-financing effect of a tax cut of about 50% on impact. Following MR and re-scaling the IRFs to a 1% of GDP increase in tax revenues ($u^\tau = 1$) would require a shock to $\varepsilon^\tau$ of about 2% of GDP, leading to their reported GDP response of 2% on impact and about 3% at the peak. Hence, the results of MR are quite sensitive to their choice of scaling. Note that due to our lower pure multipliers and elasticities, the feedback is much weaker and the choice of option (a) instead of (b) increases the impact multiplier by merely 0.1 units. Since the strength of the feedback is endogenous to the results, we prefer to compare the pure multiplier effects which amount to about 0.5 on impact (peak: 0.5) for the German case and 1 (1.6) for the US case.

(iii) Germany is a much more open economy than the US, and this should generally dampen multiplier effects through the import leakage. According to a meta analysis on fiscal multipliers by Gechert (2015), an increase in the imports-to-GDP ratio of 1 pp lowers the reported multiplier by about 0.02 units for empirical studies. With a sample average of the imports-to-GDP ratio of about 0.3 for Germany and 0.1 for the US, this would explain a difference in the multiplier of about 0.4. In combination with the scaling

Figure 4: Cumulative impulse-responses for BP (dashed blue) and MR (solid green) identification after expansionary shock of 1% of GDP to various net-revenue components (tau), log first differences, 2-SD error bands




factor, this almost aligns the findings.

## 5. Robustness

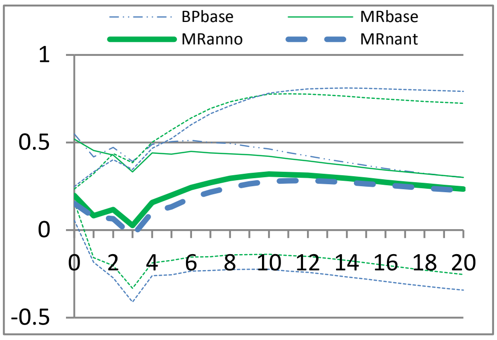
Robustness is checked in the dimensions of stochastic vs. deterministic trends, choice of identifying assumptions, model specification, and censoring of the shock series. First, we estimate a specification in log first differences to rule out an insufficient control for stochastic trends. Results are displayed in Figure 4 and contain the cumulative impulse-responses. The basic finding – multipliers from the bottom-up and top-down approaches do not differ much and rank at the lower end of the spectrum in the literature – remains robust. However, there are some relevant differences as compared to the log-level estimation: First, cumulative responses do not die out, but become permanent, in line with the original findings of MR and BP. Second, cumulative confidence bands become huge after the first quarters. While the impact estimates for the BP approach are largely unaffected, the MR impact multipliers differ somewhat in accordance with rows (6) and (7) of Table 2. The point estimates of the GDP reaction to net-revenues and taxes is now larger for the MR approach than for the BP approach. In particular, the tax multiplier increases to slightly above one for longer horizons. Note, however, that the effect is still much lower than in MR and Hayo and Uhl (2014). In contrast, social security revenue and spending multipliers for the MR approach are lower than in the baseline specification in levels and also lower than the respective BP estimates.

Further results for robustness checks are summarized in Figure 5, which contain the baseline point estimates as thin lines to foster comparison (2-SD error bands are shown for the alternative specifications). We focus on the compound net-revenue shocks and the GDP reaction for brevity here.

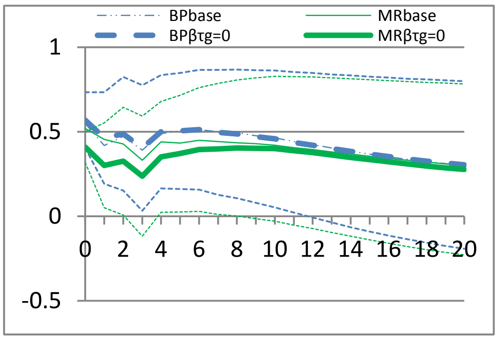
Figure 5b evaluates the choice of imposing $\beta_{g\tau} = 0$ and leaving $\beta_{\tau g}$ unrestricted vs. the opposite case. The effects on GDP are almost identical to the baseline specification.

Figure 5c presents the GDP responses for the BP and MR approaches using a more comprehensive VAR model with $X_t = \begin{bmatrix} g_t & y_t & p_t & \tau_t & i_t \end{bmatrix}'$ additionally including the log GDP deflator and a short-term nominal interest rate, like in Hayo and Uhl (2014). Identification in the MR case for the additional variables is achieved by IV estimations. For the BP case, we follow the factorization in Perotti (2005), ordering $i_t$ last and assuming $\alpha_{gp} = -0.5$ and $\alpha_{yp} = 0$. The BP case remains largely unaffected on impact but the dynamics imply a somewhat lower GDP response at longer horizons with the difference becoming significant after about 6 quarters. The MR case now produces multipliers which are statistically significantly lower than in the baseline MR specification

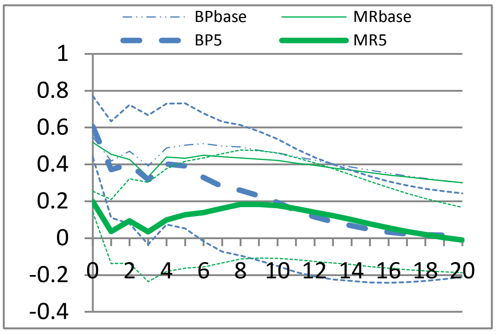
Figure 5: IRF of $y$ to change in Netrev, robustness checks with 2-SD error bands and point estimates of baseline specifications for comparison



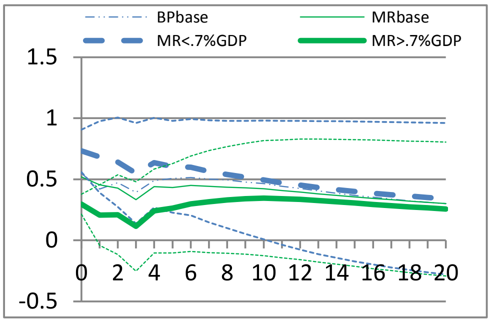

(a) Fiscal Foresight



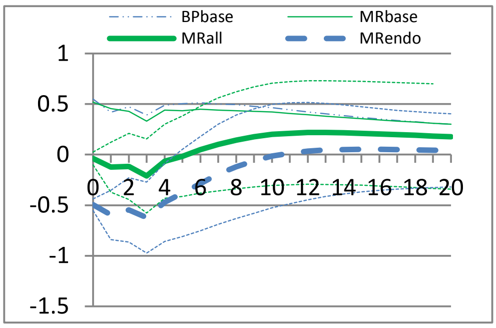

(b) $\beta_{\tau g} = 0$



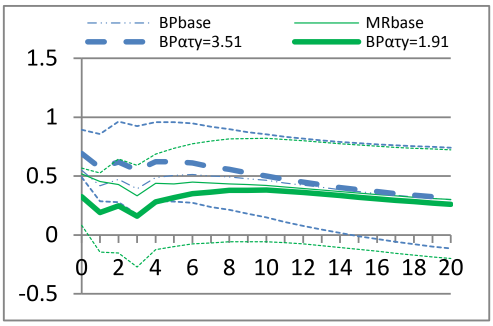

(c) 5 Variable VAR




(d) Shocks<> .7% GDP




(e) Endogenous Shocks




(f) BP-$\alpha_{\tau y}$ – sensitivity

even though the difference is not economically significant. The difference between the 5-variable BP and MR specifications is not large and statistically significant only for the first few quarters.

Figure 5d shows the GDP response, when shocks are censored to either below or above 0.7% of GDP. The GDP response to big shocks is somewhat lower, but the difference to small shocks remains significant only for the first 4 quarters.

Figure 5e shows GDP responses for estimations with the MR-approach for the endogenous shock series (MRendo) and the full discretionary series (endogenous + exogenous, MRall). Plausibly, in the case of endogenous countercyclical reactions to business cycle and other macroeconomic shocks, the IRFs for the endogenous "shocks" are downward-biased and show negative multipliers. Estimating the responses for the full shock series (MRall) yields results in between the exogenous and endogenous specifications with multipliers close to zero and insignificant.

Finally, Figure 5f includes sensitivity tests for alternative imposed elasticities of $\alpha_{\tau y}$ in the BP-approach and compares them to our baseline estimates. Since for the BP approach $\alpha_{\tau y}$ is imposed as a mere point estimate without accounting for uncertainty, we test its sensitivity by adding or subtracting the 95% confidence interval for $\alpha_{\tau y}$ as found in the MR estimation (Table 2 row (2)). Plausibly, increasing (decreasing) $\alpha_{\tau y}$ to 3.51 (1.91) comes with a higher (lower) multiplier, but the differences to the baseline case are not statistically significant.

Summing up, even though there are some level shifts of the GDP responses for alternative specifications, the differences are rarely statistically and economically significant.

## 6. Conclusion

Following the bottom-up approach of identification of exogenous fiscal policy shocks (Romer and Romer 2010), we have constructed a rich narrative data set of net-revenue shocks for Germany by (i) reconstructing the tax shock series of Hayo and Uhl (2014) and

(ii) coding an exogenous shock series for social security contributions, benefits and transfers derived from official documents of major legislative changes in pensions, health care, long-term care, unemployment insurance and basic social security. Based on quarterly data for 1974q1 to 2013q4 we have estimated the multiplier effects of changes to taxes, social security contributions and expenditures as well as net-revenues of this bottom-up identification within a proxy SVAR framework developed by Mertens and Ravn (2013). We compare them with estimates from a traditional top-down identification framework following Blanchard and Perotti (2002), and alternative bottom-up specifications (Hayo and Uhl 2014; Mertens and Ravn 2014).

Employing the bottom-up identification, we find revenue multipliers for Germany in a range of zero to slightly above one for different revenue components and specifications of our model and shock series. These estimates are much lower than in the relevant bottom-up literature so far. The multipliers as well as the implied elasticities of the revenue components do, however, square well with estimates from a top-down identification. Hence, the finding of the existing literature that bottom-up identified revenue multipliers and elasticities are much larger than conventional estimates does not seem to be generalizable. Our results may provide a step towards consensus regarding the incompatible multiplier effects on the revenue side that have been deplored in the literature so far.

## References


Baum, A. and G. B. Koester (2011), The impact of fiscal policy on economic activity over the business cycle - evidence from a threshold VAR analysis, no. 03/2011 in Deutsche Bundesbank Discussion Paper Series 1: Economic Studies.

Blanchard, O. and R. Perotti (2002), An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output, *Quarterly Journal of Economics 117*(4), 1329–1368.

Bundesministerium für Arbeit und Soziales (2011), Gesetzestexte zur Sozialgeschichte 1881 –2010.

Caldara, D. and C. Kamps (2012), The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers, no. 2012-20 in Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board.

Caldara, D. and C. Kamps (2013), The Identification of Fiscal Multipliers and Automatic Stabilizers in SVARs, Mimeo.

Cloyne, J. (2011), What are the effects of tax changes in the United Kingdom? New evidence from a narrative evaluation, no. 3433 in CESifo working paper.

Cloyne, J. (2013), Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom, *American Economic Review 103*(4), 1507–1528.

Favero, C. and F. Giavazzi (2012), Measuring Tax Multipliers: The Narrative Method in Fiscal VARs, *American Economic Journal: Economic Policy 4*(2), 69–94.

Gechert, S. (2015), What fiscal policy is most effective? A meta-regression analysis, *Oxford Economic Papers 67*(3), 553–580.

Gechert, S., C. Paetz and P. Villanueva (2016), A Narrative Account of Legislated Social Security Changes for Germany, no. 170 in IMK working paper.

German Statutory Pension Insurance (2011), Rentenversicherung in Zeitreihen.

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

Lütkepohl, H. (2006), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin u. a.

Mertens, K. and M. O. Ravn (2010), Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach, *Economic Journal 120*(544), 393–413.

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.

Mertens, K. and M. O. Ravn (2014), A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers, *Journal of Monetary Economics 68*(S), S1–S19.

Perotti, R. (2005), Estimating the effects of fiscal policy in OECD countries, no. 4842 in CEPR Discussion Papers.

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, URL http://www.nber.org/papers/w16786.

Price, R., T.-T. Dang and Y. Guillemette (2014), New Tax and Expenditure Elasticity Estimates for EU Budget Surveillance, no. 1174 in OECD Economics Department Working Papers.

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*(3), 763–801.

Romer, C. D. and D. H. Romer (2014), Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Changes, 1952-1991, no. 20087 in NBER working paper.

Steffen, J. (2013), Sozialpolitische Chronik: Die wesentlichen änderungen in der Arbeitslosen-, Renten-, Kranken- und Pflegeversicherung sowie bei der Sozialhilfe (HLU) und der Grundsicherung für Arbeitsuchende - von den siebziger Jahren bis heute, no. [15/01/2014] in http://www.portal-sozialpolitik.de.

Tenhofen, J., G. B. Wolff and K. H. Heppke-Falk (2010), The macroeconomic effects of exogenous fiscal policy shocks in Germany: a disaggregated SVAR analysis, *Jahrbücher für Nationalökonomie und Statistik 230*(3), 328–355.

Uhl, M. (2013), A history of tax legislation in the Federal Republic of Germany, no. 11-2013 in MAGKS Joint Discussion Paper Series in Economics.

## A. Appendix

We construct the imposed elasticities for the top-down approach on the basis of (Perotti 2005). However, we do not use Perotti's auxiliary regressions to estimate the impact of GDP changes to tax bases, but stick closely to the full OECD figures. Income tax elasticity to its base consists of the weighted elasticity of its components, which is: (i) Taxes on earnings to earnings as given by the latest OECD estimate of 1.9 (Price et al. 2014), leading to an earnings tax to output gap elasticity of 1.32, with a weight of 71.25% in income taxes. (ii) Self-employed income (weight: 21.25%), and (iii) capital income (weight: 7.5%), should both have a zero contemporaneous elasticity to their base, according to the tax code, due to collection lags and our use of cash data for the fiscal variables.[4] Overall, the income tax elasticity would be 0.94 in our case.

The overall tax elasticity consists of the income tax elasticity (weight: 39.83%), corporate tax elasticity (weight: 11.86%) and indirect tax elasticity (weight: 48.31%). Perotti (2005) argues that corporate tax payments in a given quarter are based on previous

[4]Capital income taxation has changed in 2009. Since then, dividends are subject to a withholding tax. However, this system is in place since 2009 only, and does not apply to the lion's share of our sample.

years' estimates and should thus have a zero contemporaneous elasticity. However, corporate tax and trade tax advances in Germany are corrected for current performance within the quarter. Thus, a larger than zero elasticity is warranted and we make use of the OECD estimate of 1.91, deviating from Perotti's choice, even though this has only a minor effect on the overall tax elasticity given the low weight. With respect to indirect taxes, in accordance with Perotti (2005), we rely on the unit elasticity that the OECD prefers. We calculate an overall tax elasticity of 1.08.

With respect to social security contributions, we follow on the OECD measure of the contributions-to-output-gap elasticity of 0.6. Social security expenditures, including transfers are also partly elastic to the cycle, in particular unemployment benefits (-3.3, weight: 10.47%) and earnings-related benefits (-0.64, weight: 23.49%), with the remainder assumed inelastic. Hence, the average elasticity of social spending amounts to -0.50.

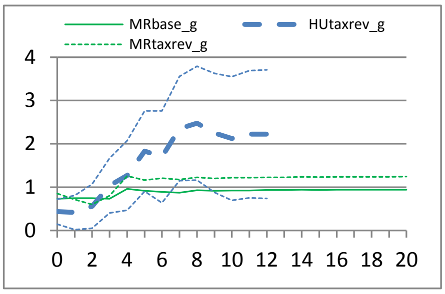
Combining these measures, net revenues, as a mixture of taxes (weight: 138.35%), social security contributions (weight: 103.56%), social security expenditures (weight: -122.44%) and inelastic interest payments (weight: -17.5%), have an elasticity of 2.71 with respect to GDP, which is much higher than the value of 0.92 employed by Perotti (2005). Using Perotti's low value, we also find negative net-revenue multipliers on impact for Germany of about -0.15, that turn significantly positive only after the thirteenth quarter.

## B. Appendix

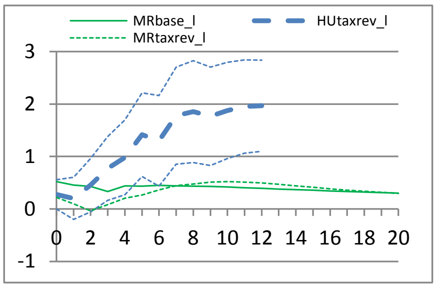
In this appendix we reproduce the results of Hayo and Uhl (2014) and Mertens and Ravn (2014).

Despite some minor deviations with respect to the sample size, the availability of endogenous variables, and uncertainty regarding the definition of their dummy variables, we can reproduce the Hayo and Uhl (2014) results quite closely. The model reads as

Figure 6: Dynamic real GDP per capita response after an expansionary shock to Taxes of 1% GDP, bottom-up identification




(a) First differences




(b) Levels

Note that HU use 1-SD error bands. We follow them here, as for their framework 2-SD confidence bands are rather wide.

follows:

$$\mathbf{\Gamma}(L)X_t = v + \boldsymbol{\lambda}(L)m_t + w_t \tag{11}$$

$$X_t = \begin{bmatrix} g_t & y_t & p_t & \tau_t & i_t \end{bmatrix}' \tag{12}$$

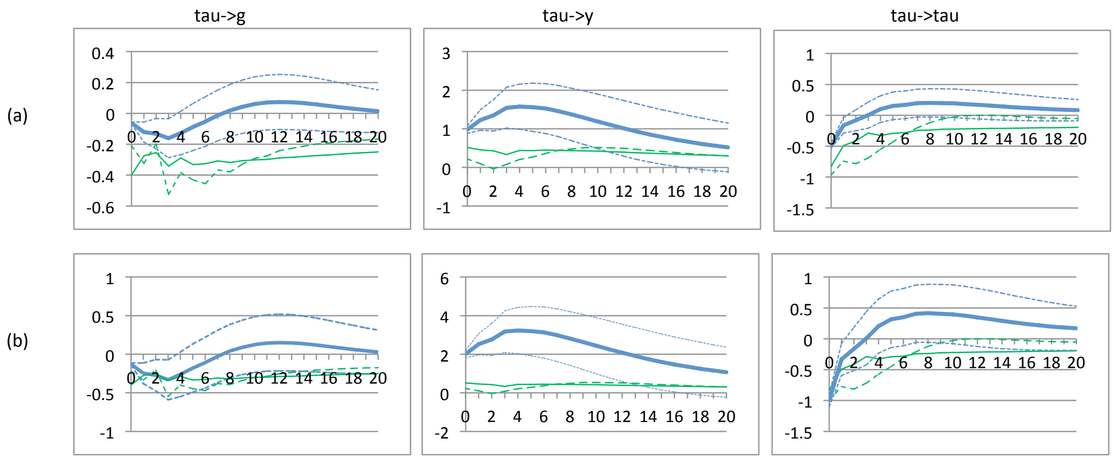
Again, $\Gamma(L)X_t$ is a 4th-order lag polynomial of the (lagged) endogenous variables $X_t$ and their coefficients as described in Section 3. In line with Hayo and Uhl (2014), the lag length of the endogenous variables is set to four.

On the right hand side, $w_t$ are reduced form residuals, $v$ again includes a constant, a re-unification dummy and a financial crisis dummy. $\boldsymbol{\lambda}(L)m_t$ is an 8th-order lag polynomial containing the (lagged) exogenous tax shock series and its coefficients. Again, eight lags are in line with the baseline specification in Hayo and Uhl (2014), but note that the resulting dynamic multipliers are quite sensitive to this choice.

Figure 6 shows the dynamic multiplier effect of $y$ to a shock in taxes for the Hayo and Uhl (2014) framework. We display results for specifications in first differences and levels. For comparison, we add the point estimates of our proxy SVAR tax and net-revenue multipliers.

Figure 7 reproduces the MR US estimates (sample spans from 1950q1 to 2006q4)

Figure 7: Impulse-responses for MR US estimates (solid blue) and MR Germany net revenues (solid green) and taxes (dashed green) to an expansionary shock of 1% of GDP in effective tax revenues (option (a)) or prospective tax liabilities (option (b)), log levels, 2-SD error bands




based on option (a) and (b) as described in Section 3. For comparison, we add the point estimates of tax and net-revenue multipliers for our German sample. For the US sample the feedback through the multiplier effect itself and the budget elasticity is quite strong such that re-scaling the shocks to 1% of GDP in effective revenues about doubles the GDP reaction.

## Impressum

**Publisher:** Hans-Böckler-Stiftung, Hans-Böckler-Str. 39, 40476 Düsseldorf, Germany
**Phone:** +49-211-7778-331, IMK@boeckler.de, http://www.imk-boeckler.de

**IMK Working Paper is an online publication series available at:**
http://www.boeckler.de/imk_5016.htm

**ISSN:** 1861-2199