The Macroeconomic Effects of Social Security Contributions and Benefits – Evidence from Germany

Sebastian Gechert, Christoph Paetz and Paloma Villanueva*

September 24, 2018[†]

Abstract. We construct a novel narrative dataset of legislated social security shocks for Germany. The dataset covers major changes in benefits and social security contributions for pensions, health care, long-term care and unemployment insurance on the German federal level over 1974q1 to 2013q4. We estimate the multiplier effects of the narratively identified shocks in a proxy SVAR framework and compare them with estimates found via traditional top-down identification. We find social security multipliers for Germany between 0.5 and 1.5. The GDP response to changes in contributions fades relatively quickly, while expenditures imply somewhat more persistent and slightly stronger effects.

Keywords. Narrative Approach; Fiscal Multiplier; Social Security

JEL classification. E62, H20, H30

*Gechert: Macroeconomic Policy Institute (IMK), Hans-Boeckler-Strasse 39, 40476 Düsseldorf, Germany. Email: sebastian-gechert@boeckler.de. Tel: +49 211 7778 306. Paetz: Macroeconomic Policy Institute (IMK), Email: christoph-paetz@boeckler.de. Villanueva: Complutense Institute for International Studies (ICEI). Email: pvillanuevacortes@gmail.com. Declarations of interest: none

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

1 Introduction

In recent years there has been a tremendous surge in the literature on the size of fiscal multipliers. While many papers have focussed on the effects of federal and local public procurement, employment and investment spending and tax shocks, the impact of changes in social security contributions and benefits has received only little attention. This seems surprising given the fact that social security systems have grown substantially in OECD countries after the Second World War and account for about half of the overall budget in countries like Germany.

Empirical evidence is very much focused on US data: Using monthly data from 1965 to 1985 for the US, Wilcox (1989) provides evidence that anticipated positive shocks to social security benefits have significant positive short-term effects on consumption expenditure, thus rejecting the zero-response predicted by the life-cycle hypothesis. Romer and Romer (2016), exploiting narratively identified shifts in US social security spending, investigate the impact of benefit increases on private consumption expenses (where they find a substantial short-run effect) and industrial production and unemployment (where they find no significant effect).

Since headline macroeconomic budgetary figures are prone to considerable endogeneity with respect to cyclical fluctuations, shifts of the revenue and transfer 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 has centered around identification strategies towards truly exogenous variations in fiscal policy, excluding automatic stabilisers as well as discretionary, yet endogenous reactions to business cycle events and other macroeconomic shocks that could distort the measured multiplier effects. The main variants in the literature are the narrative approach of Romer and Romer (2010) (RR henceforth) and the Blanchard and Perotti (2002) (BP henceforth) structural VAR (SVAR) approach. RR identify size, timing and motivation of legislative changes with a relevant budgetary impact via legislative texts and other historic documents and create a shock series from the bottom-up. BP cyclically adjust headline budgetary figures via imposing estimated elasticities of budgetary components and lags in the reactions of budget planners to output shocks. While narrative approaches have been considered to give a more reliable and comprehensible identification, they may be more prone to measurement error and incomplete information. Thus, there is still no consensus about the least biased strategy (Caldara and Kamps 2017).

This paper provides new evidence on the macroeconomic effects of social security shocks for Germany. As compared to Wilcox (1989) and Romer and Romer (2016), we investigate the overall output effects of both benefits and contributions to the social security system. We are not aware of any such comparison of multiplier effects of benefits and contributions in the macroeconomic literature so far. A likely channel driving possible differences between the two are distributional effects, since changes in benefits tend to affect the poor while contributions are largely paid by the non-poor.

Moreover, given the difficulties in identification, we follow a two-pronged strategy: First, following the narrative approach and 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 novel time series of legislated social security shocks for Germany, both for expenditures and revenues. The dataset covers major changes in benefits 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 1974 to 2013. We then feed these narrative shocks as proxies for the latent structural shocks into a proxy SVAR model à la Mertens and Ravn (2014) (MR henceforth), including the social security system, general government spending and GDP. From this, we estimate the respective social security benefits and contribution multipliers. Second, we compare these estimates with those from the BP identification based on the same macroeconomic variables and using the latest official figures of budget elasticities as identifying restrictions.

The narrative MR proxy SVAR estimation yields statistically significant social security revenue multipliers of about 0.8 on impact. The effect continuously peters out and disappears after ten quarters. The impact multiplier for benefits is slightly higher with 0.9, decreases slowly and approaches 0.3 at a 20-quarter time horizon after the shock. These somewhat higher multiplier effects of benefits can be rationalized by heterogeneous marginal propensities to consume (MPC): since the social security system redistributes from richer to poorer households and the latter likely have a higher MPC (Jappelli and Pistaferri 2014), the net effect on aggregate demand would turn out positive.

For both contributions and benefits, results are close and not significantly different to estimates from the BP SVAR approach. Thus, different approaches to identification do not seem to affect our results. The huge discrepancy of tax multipliers using either the narrative or the cyclical adjustment approach that is discussed in MR and Caldara and Kamps (2017) does not carry over to our social security data.

In conclusion, we find that expansionary social security changes have a positive short-to-medium-term impact on GDP for Germany, somewhere in the middle of the range of multipliers in the literature (Gechert 2015). Given the size of the social security system and its frequent substantial law changes, they are likely as relevant for macroeconomic dynamics as are those to the tax system and general government spending, making them a relevant subject for future research.

The remainder of the paper is organized as follows. In Section 2, we describe the construction of the narrative shock series and examine their properties. Section 3 introduces the econometric framework and the alternative approaches to identification. In Section 4, we present our findings regarding the multiplier effects of both the MR and BP specification. 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 social security contributions and benefits following the bottom-up approach for Germany. A detailed description of the construction of our social security shock series law by law can be found in the companion paper (Gechert et al. 2018).

In contrast to tax laws, 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 narrative 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 RR we assign to each law an exogenous or endogenous underlying motivation. However, we re-evaluate these information with external sources like newspapers or major economic events. We classify those measures as endogenous, which are either (a) countercyclical or procyclical policies, (b) reactions to other macroeconomic shocks (like financial crises, oil price shocks, etc.) or (c) 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). Note that a procyclical motivation has not been identified in the original RR paper. However, we noted several such cases for German legislations where budget deficits as a result of recessions have been answered by increases in contribution rates and vice versa. Procyclically motivated interventions mainly occur in situations where there is an immediate concern about a deficit in the social security budget.

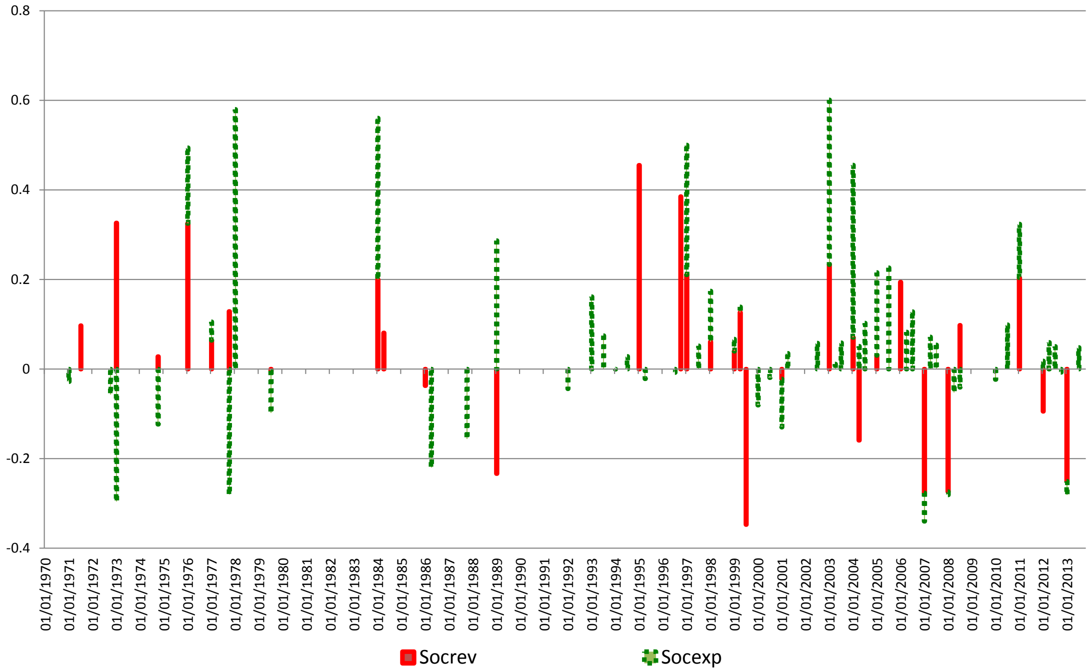
Refraining to consider these endogenous measures in the shock series should rule out likely biases from omitted variables or reverse causality. The relevant exogenous changes that lend themselves to a causal interpretation with respect to short-run multiplier effects are those that are motivated by (d) attempts to long-term budgetary consolidation, as they tackle inherited debt unrelated to current circumstances, (e) structural or ideological reasons for example to increase potential output growth or (f) rulings of the court that are usually related to cases with long decision lags, again unrelated to current circumstances. See also Romer and Romer (2010) for further discussion of these categories.

(ii) From the information of the law, we are able to detect the timing of implementation of a measure in order to determine the quarter of the shock in our data set. 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 some rare cases we lacked sufficient information on prospective implementation dates from the legislative texts or found their timing inconsistent with budgetary data from the financial statistics of the Bundesbank (*Finanzstatistik*). In such cases we used the latter to date the shocks. In the event that measures are of a temporary nature, the date of its expected expiration is recorded as well and provides the timing where we set the respective counter-shock (of the same size but opposite sign). When temporary measures are prolonged, a new shock with the new expiration date is included. Announcement effects are coded uniformly for all measures of a law. The publication date of the draft lends itself to determining the announcement date, as it provides detailed information and usually comes with newspaper coverage. If there is additional information that point to an earlier or later announcement date, we take them into account. Looking at the shock series through the lens of announcement dates will switch off any temporary measures since both the shock and its reversal will have the same announcement date.

(iii) The size of the shock and the economic relevance of each law is determined by its total prospective full-year impact (*Volle Jahreswirkung*), which is usually given in the draft of the law, divided by annual nominal GDP in the year of the shock. The total prospective full-year impact is defined as the annualized financial impact after full implementation, but assuming no change in the respective base of contributors or beneficiaries. This is a feature, since it clearly separates the size of the pure shock from these endogenous reactions by agents, which are reflected in the dynamic response of headline budgetary figures and GDP. We include all laws in the shock series with a prospective full-year impact above or just slightly below 0.1% of annual nominal GDP. 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. Since the calculation of the budgetary impact as given in the draft legislation may be prone to forecast errors or political bias, using the proxy SVAR approach in our econometric specification is particularly appealing as it can cope with measurement error.

Figure 1 summarizes our exogenous shock series at implementation dates from 1970 to 2013. For social security contributions (Socrev) the mean is positive but very low with 0.009 (standard deviation: 0.08) and for social security benefits (Socexp) 0.015 (sd:

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



0.09), positive and close to zero as well. Volatility is low for both social security shock series.

Figure 2 includes those shocks for Socrev and Socexp which are endogenously motivated. The endogenous series of Socrev also has a mean close to zero with 0.008 (sd: 0.07), Socexp -0.005 (sd: 0.08). Endogenous policy actions to the social security system were concentrated at the beginning of the 1980s, the mid 1990s and after the financial crises.

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 (four

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

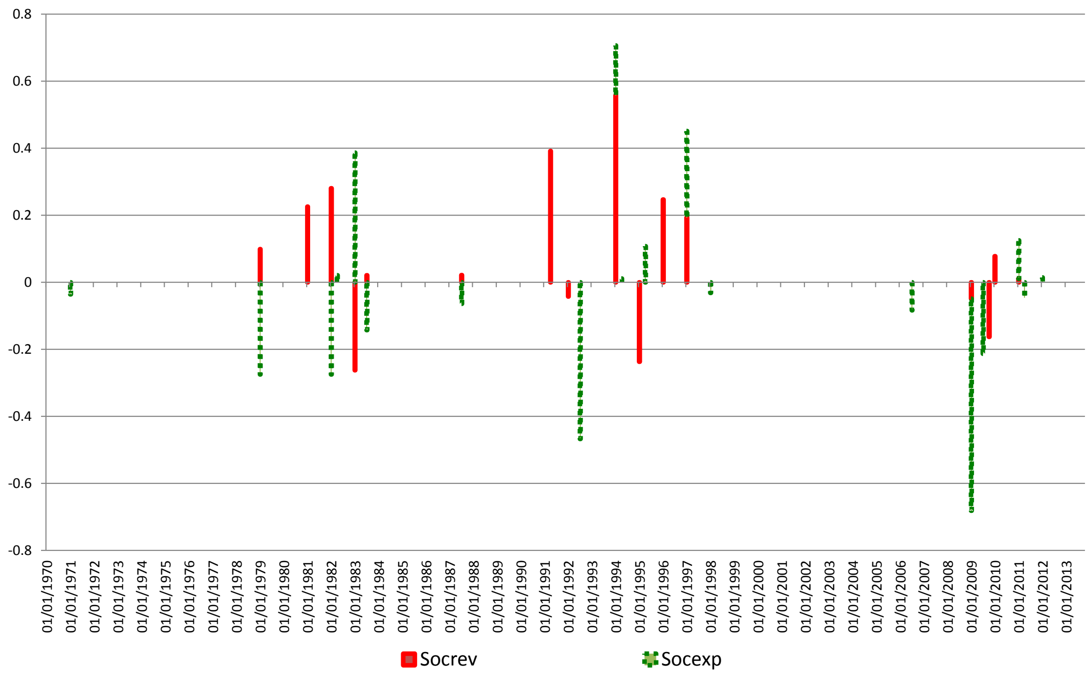


Table 1: Predictability of the shock series – Granger causality tests (based on lags 1 through 4 of growth rates of GDP, government expenditures and the respective social security component.)

		Socrev	Socexp
Exo	χ^2	7.844	14.700
	$p(\chi^2)$	0.797	0.258
Endo	χ^2	22.378	17.695
	$p(\chi^2)$	0.033	0.125

lags of GDP, government expenditures and the respective headline time series for social security contributions and benefits). 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.

Dealing with the issue of fiscal foresight in more detail, we discuss the case of legislations, whose implementation follows swift after their publication such that anticipatory effects can be largely ruled out in Section 5.

3 Model, Data and Identification

The VAR model for both identification approaches can be summarized as follows:

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

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

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

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 Γ. For our baseline estimation, all variables are in log-levels. The choice of a 4th-order lag polynomial is based on the quarterly structure of our dataset and has become a standard choice in the fiscal VAR literature for quarterly data.

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 social security contributions or benefits, respectively (τ_t). We also 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 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.

v contains a constant, a linear time trend, a re-unification dummy (1991q1-2013q4 in levels and 1991q1-q4 in first differences) and a financial crisis dummy (2009q1). u_t is the $K \times 1$ vector of reduced-form disturbances, while ε_t contains the $K \times 1$ structural-form shocks that are to be identified by either the MR or BP 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 ε_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 Σ_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 $\boldsymbol{B}$.

3.1 Mertens-Ravn identification

Following the narrative approach, identification is achieved by recording exogenous changes to social security legislation, determining their motivation, timing and impact on revenues or transfers, 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 RR 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).

$$\boldsymbol{\Gamma}(L)X_t = v + \boldsymbol{\lambda}(L)m_t + w_t \tag{6}$$

They then proceed by estimating dynamic multiplier functions of X_t to a shock in m_t. However, these dynamic multiplier functions are not 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.[1] 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^τ using the shock series m_t as the instrument.

$$\hat{u}_t^i = \mu^i + \alpha_{i\tau}^{IV} \bar{u}_t^\tau + \zeta_t^i \tag{7}$$

$$\hat{u}_t^\tau = \mu^\tau + \gamma m_t + \zeta_t^\tau = \bar{u}^\tau + \zeta_t^\tau \tag{8}$$

Table 2 shows tests of the relevance and reliability of the instrument in the 2SLS-regression (7) both for the specification of the SVAR in log-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 ε_t^τ is derived by regressing the estimated structural shocks $\hat{\varepsilon}_t^\tau$ on the non-zero observations of m_t. The resulting R^2 statistic should asymptotically signal the reliability of the instrument (Mertens and Ravn 2013). A high R^2 can be interpreted in the way that an identified narrative shock m_t in a certain quarter can explain a substantial part of the contemporaneous variation in headline social security spending and revenue figures. For both components of social security, the instrument seems highly relevant and m_t shocks have reasonable predictive power for $\hat{\varepsilon}_t^\tau$, despite the limited information in the narrative sources and possible measurement error.

(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

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

Table 2: Relevance and reliability of the instrument

		Socrev	Socexp
L	$F(u_t^\tau, m_t)$	15.454	17.154
	$p(F)$	0.000	0.000
	$R^2(\varepsilon_t^\tau, m_t)$	0.249	0.319
G	$F(u_t^\tau, m_t)$	16.720	16.397
	$p(F)$	0.000	0.000
	$R^2(\varepsilon_t^\tau, m_t)$	0.313	0.226

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\bar{\alpha}_{gy} & -\bar{\alpha}_{g\tau} \\ -\alpha_{yg} & 1 & -\bar{\alpha}_{y\tau} \\ -\bar{\alpha}_{\tau g} & -\alpha_{\tau y} & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} \beta_{gg} & 0 & \bar{\beta}_{g\tau} \\ 0 & \beta_{yy} & 0 \\ \beta_{\tau g} & 0 & \beta_{\tau\tau} \end{bmatrix} \tag{9}$$

where $\bar{(\cdot)}$ denotes a restricted parameter. Imposing the following restrictions will be sufficient for a just-identified model: Leaving $\beta_{\tau g}$ unrestricted and setting $\beta_{g\tau} = 0$ implies that in the process of budget planning, direct government spending decisions are taken prior to social security budget decisions. We show the robustness of this choice for our sample in Section 5. Moreover, endogenous social security contributions and benefits are assumed not to be driven by direct government spending within the same quarter ($\alpha_{\tau g} = 0$) apart from its indirect influence via affecting output ($\alpha_{yg}\alpha_{\tau y}$). Government direct spending (excluding transfers and interest) is assumed to be inelastic to GDP within a quarter ($\alpha_{gy} = 0$). These restrictions are all in line with Caldara and Kamps (2008). The $\alpha_{y\tau}$ elasticities can be transformed into impact multipliers by re-scaling the usual 1-SD shocks to 1% of GDP changes with the sample-average ratio of τ/y (in linear levels).

The main distinction between MR and BP is that in MR, $\alpha_{i\tau}$ are determined by the IV regression, while leaving the critical elasticity $\alpha_{\tau y}$ of contributions or benefits to changes in GDP unrestricted. The opposite applies to the BP case as discussed in the following section.

3.2 Blanchard-Perotti identification

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

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

Again we restrict $\alpha_{gy} = \alpha_{\tau g} = \beta_{g\tau} = 0$. Moreover, government direct spending is assumed to be inelastic to social security changes within a quarter ($\alpha_{g\tau} = 0$). The crucial assumption for estimating social security multipliers with the BP approach concerns the elasticity of contributions and benefits to GDP $\alpha_{\tau y}$. We determine $\alpha_{\tau y}$ for our different categories based on the latest OECD estimates (Price et al. 2014). With respect to social security contributions, we follow the OECD measure of the contributions-to-output-gap elasticity of 0.60. Social security expenditures, including transfers are also partly elastic to the cycle, in particular unemployment benefits (-3.3, with a share of 10.47% in total benefits) and earnings-related benefits (-0.64, share: 23.49%), with the remainder assumed inelastic. Hence, the weighted average elasticity of social spending amounts to -0.50.

Caldara and Kamps (2017) 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 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. Comparing the values with the estimates from the MR approach provides a useful test as to whether the restrictions are valid.

Table 3: Elasticities imposed and estimated for the BP and MR models in levels (L) and growth rates (G)

		Socrev	Socexp
$\alpha_{\tau y}$			
(1)	BP imposed L+G	0.60	-0.50
(2)	MR implied L	0.74 (0.43, 1.04)	-0.74 (-1.14, -0.34)
(3)	MR implied G	0.46 (0.14, 0.78)	-0.62 (-1.04, -0.2)
$\alpha_{y\tau}$			
(4)	BP implied L	-0.09 (-0.18, 0)	0.15 (0.07, 0.23)
(5)	BP implied G	-0.09 (-0.18, -0.01)	0.10 (0.03, 0.17)
(6)	MR imposed L	-0.13	0.20
(7)	MR imposed G	-0.06	0.12

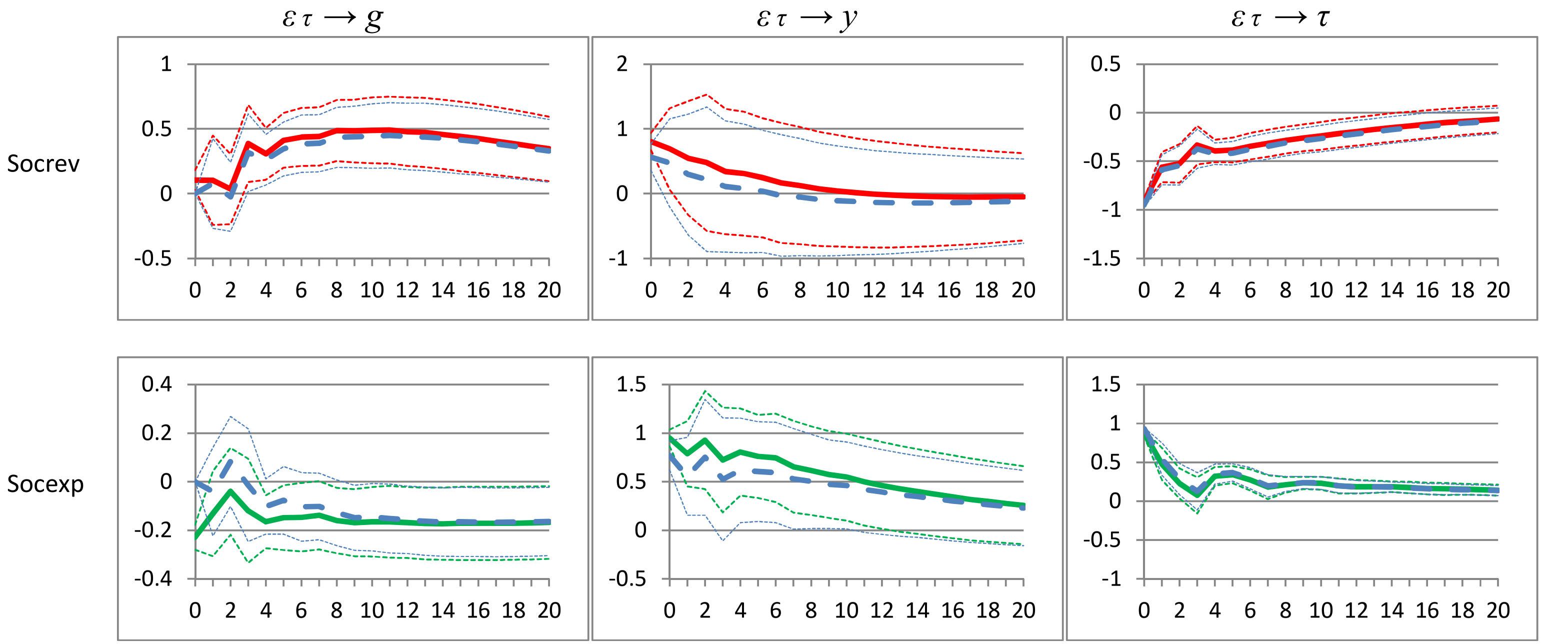
95% confidence bounds for implied elasticities in parentheses.

Setting the $\alpha_{\tau y}$ value has the advantage that the contemporaneous reaction of GDP to changes in social security revenues and expenditures $\alpha_{y\tau}$ can be left unrestricted and be determined by the data. Rows (4) and (5) of Table 3 list the implied elasticities of y to a change in τ for our estimation of the BP SVAR both in log-levels (L) and growth rates (G).

For comparison, the imposed and implied elasticities for the BP and MR approach can be found in Table 3. The imposed budget elasticities ($\alpha_{\tau y}$) used in the BP case are given in row (1). One can also calculate the implied budget elasticities out of the MR approach. They embrace the BP figures within the 95% confidence interval. According to these figures, the German social security system is slightly progressive in sum. That is, an increase in GDP by 1% will improve the budget of the social security system by 1.1% (BP) to 1.5% (MR).

Turning to the elasticities of output to social security shocks ($\alpha_{y\tau}$), rows (4) and (5) display the implied elasticities that turn out of the BP estimation. Rows (6) and (7) list the imposed elasticities of y to a change in τ from the MR proxy estimation both in levels (L) and growth rates (G). As a mirror image from the above, they also fall within the 95% confidence interval of the BP figures.

Figure 3: Impulse-responses for MR (solid red or green) and BP (dashed blue) identification after expansionary shock to social security contributions or benefits (tau) sized to 1% of GDP, log levels, 2-SE confidence bands



4 Results

We now estimate the responses of the endogenous variables to an expansionary shock ($-\varepsilon_t^{\tau}$) to the respective series, that is, either a relief in social security contributions (Socrev), or an increase in benefits and transfers (Socexp). Shocks are sized to 1% of GDP of prospective statutory revenues or expenditures. The error bands are 2-SE centered confidence intervals from a recursive wild bootstrap (Gonçalvez and Kilian 2004).[2] Figure 3 shows the IRFs of our baseline specification in levels.

Using both the MR and BP methodology and identification, we find plausible multiplier effects on y slightly below one on impact and significantly different from zero for both social security contributions and expenditures. The MR impact multipliers are slightly higher than the BP ones, but the difference is not statistically significant. 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 3. The GDP response to changes in social security contributions peters out relatively quickly, becoming insignificant at the 95% level soon after impact and approaching zero after about 2 years. Social security expenditures imply only a slightly higher impact multiplier, but the GDP effects are much more persistent, staying significant for about 3 years and with a point estimate 0.3pp above baseline after 5 years.

These effects are very much in line with the multipliers from the broader literature on tax changes and other government spending. The meta analysis of Gechert (2015) reports an average impact multiplier of about 0.8 for general public spending and 0.5 for

[2] The error bands are based on 10000 replications. The bootstrapped standard errors are almost identical to those from analytic methods for the log-level estimates, but smaller than the analytic ones for the cumulative impulse-responses of the estimates in first differences. Note that we parsimoniously reduce the number of replications to 1000 for further robustness tests. The confidence intervals of our baseline results are narrow compared to e.g. Romer and Romer (2010), but in line with Mertens and Ravn (2014). This might be specifity of the proxy SVAR approach. We obtain large error bands, comparable to Romer and Romer (2010) when we use an ARDL approach as can be seen further below in the robustness section 5. Thus the narrow standard errors of the proxy SVAR may be taken with a grain of salt.

taxes. Comparing these findings with Romer and Romer (2016), the effects for benefits seem to be comparable on impact, but are much more persistent in our case. Moreover, Romer and Romer (2016) do only find effects on consumption that do not spill over to employment or industrial production. For the German case, there is a non-negligible impact on output in general.

How can we make sense of the differences between benefits and contributions? Jappelli and Pistaferri (2014) show, based on Italian survey data, that marginal propensities to consume are higher for low cash-on-hand percentiles. Given that benefits are likely pro-poor while contributions are paid by middle and upper income classes, it seems plausible that benefit shocks have a stronger aggregate demand effect. Moreover, some benefits are in-kind and will have a direct GDP effect.

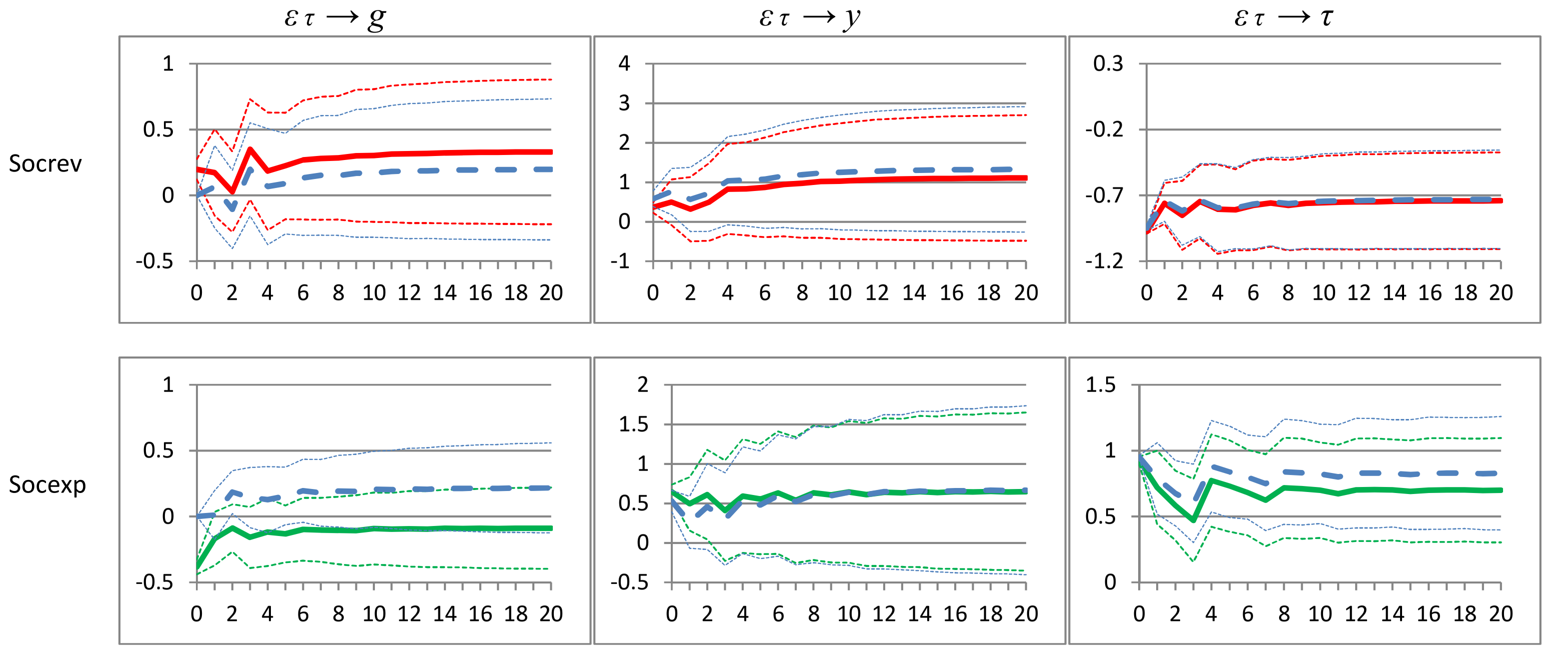
The response of τ itself is also temporary and dies out almost entirely over the 20-quarter horizon. This may partly be attributed to automatic stabilizers: the GDP expansion endogenously lowers expenses and raises revenues over the course of time. Moreover, ε_τ shocks may not be fully permanent in levels themselves in a growing economy. The transitory dynamics of τ may also explain the temporary nature of the GDP response.

The reaction of non-transfer government spending (g) is close to zero on impact but becomes positive and just statistically significant after a cut in social security revenues.[3] This might be interpreted as an endogenous reaction of government spending to increasing GDP. However, g reacts slightly negative (and just significantly so) after an increase in social security spending. Assuming that there are constraints to the overall level of public spending, there could be a substitution effect away from non-transfer spending.

In order to rule out an insufficient control for stochastic trends, we also estimate a specification in log first differences. Results are displayed in Figure 4 and contain the cumulative impulse-responses.

[3]Note that the initial response is restricted to zero in the BP specification.

Figure 4: Cumulative impulse-responses for MR (solid red or green) and BP (dashed blue) identification after expansionary shock to social security contributions or benefits (tau) sized to 1% of GDP, log first differences, 2-SE confidence bands



The basic finding – multipliers from the MR and BP approaches do not differ much and are close to or below one – remains intact. However, there are some relevant differences as compared to the log-level estimation: First, impact multipliers for the MR approach are now somewhat lower than in the baseline specification in levels (0.36 for revenues and 0.65 for expenditures), while there is almost no change for the BP case in accordance with rows (4) through (7) of Table 3. Second, cumulative IRFs do not die out, but become permanent, contrary to the transitory dynamics in the level specification. This pattern is in line with the original findings of MR and BP (for their tax multipliers) and likely relates to the poorer dynamics imposed by first-differencing.[4] Third, the GDP response is higher than in the levels specification, in particular for social security revenues. Fourth, cumulative confidence bands become much larger after the first quarters. This lower precision of the estimates is quite frequent in studies employing dynamic multiplier estimates, as seen in RR and Hayo and Uhl (2014). Hence our findings for the first-difference specification suggest caution regarding the much narrower confidence intervals in the level specifications.

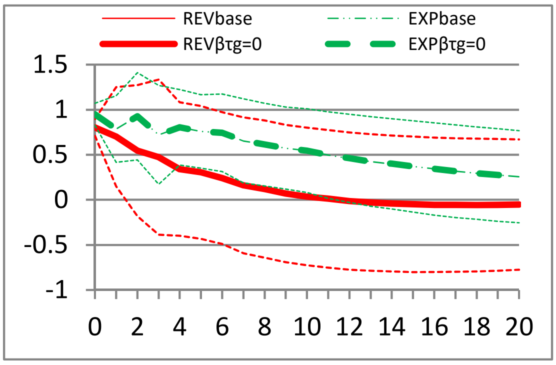
5 Robustness

Robustness is checked in the dimensions of choice of identifying assumptions, model specification, censoring of the shock series, and splitting the sample into different time periods.

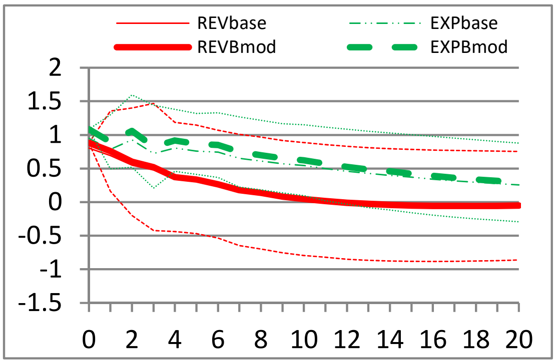
Results for robustness checks are summarized in Figure 5 and 6, which contain the baseline point estimates as thin lines to foster comparison. Figure 5 concentrates on robustness for model uncertainty and Figure 6 checks different specifications regarding the shock series and sample. We focus on the GDP reaction for brevity here.

[4]The non-cumulative IRFs fall to zero without delving into negative territory. Therefore the cumulative IRF converges to a positive level after some quarters. Cumulative multiplier effects ($\sum_h \Delta y_h / \sum_h \Delta \tau_h$) calculated over the whole 20-quarter horizon for the first-difference specification (REV: 1.2, EXP: 0.9) are close to the impact multipliers for the level specification. This shows that the overall findings are similar even though the dynamics are somewhat different.

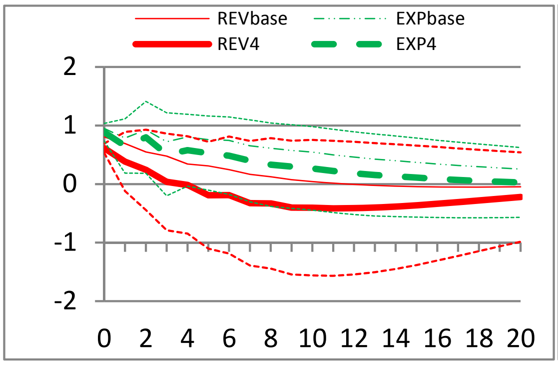
Figure 5: IRF of y after expansionary shock of 1% of GDP in Socrev or Socexp, model uncertainty robustness checks with point estimates of baseline specifications for comparison



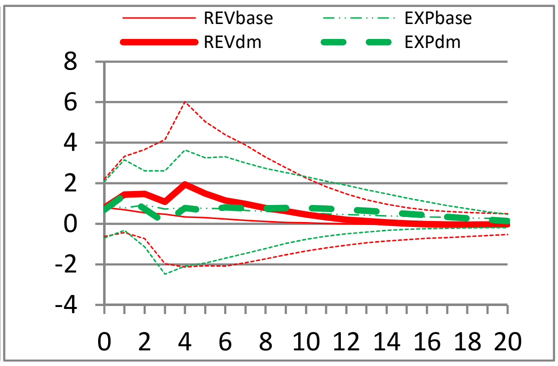
(a) $\beta_{\tau g} = 0$



(b) B Model

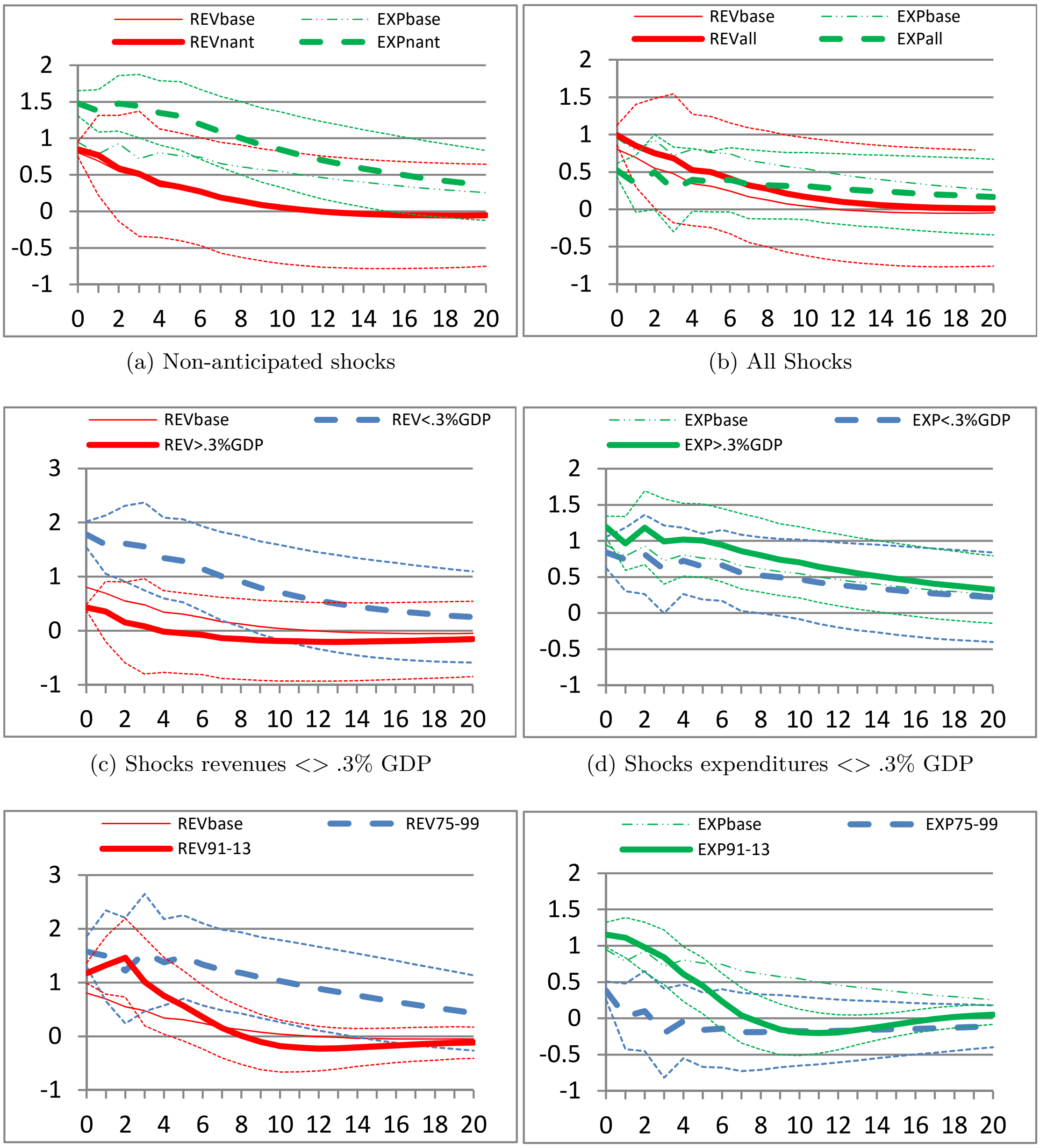


(c) 4 Variable VAR



(d) Dynamic Multipliers

Figure 6: IRF of y after expansionary shock of 1% of GDP in Socrev or Socexp, sample uncertainty robustness checks with point estimates of baseline specifications for comparison



(a) Non-anticipated shocks

(b) All Shocks

(c) Shocks revenues <> .3% GDP

(d) Shocks expenditures <> .3% GDP

(e) Sample Split Socrev

(f) Sample Split Socexp

Figure 5a evaluates the choice of imposing $\beta_{g\tau} = 0$ and leaving $\beta_{\tau g}$ unrestricted vs. the opposite case. Results are practically identical such that the IRFs cover their baseline counterparts. Results are only shown for the MR case, but apply to the BP case as well.

Figure 5b covers the case of using a $\mathfrak{B}$ model of factorization, like MR used in their original contribution, instead of our baseline AB factorization. The difference boils down to the question as to whether the shock size accounts for contemporaneous feedback via automatic stabilizers: Either it can be interpreted (I) as an increase in projected statutory contributions or benefits of 1% of GDP ($\varepsilon^\tau = 1$), *excluding* the initial feedback on its base or (II) as an increase of effective contributions or benefits of 1% of GDP *including* the feedback ($u^\tau = 1$). In their paper, MR follow option (II) 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 (ε^τ) 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 (I) and (II). By the IV estimation, we determine the $\mathbf{A}_\tau$ column vector. At the expense of setting the additional identifying restrictions in order to disentangle the $\mathbf{A}$ and $\mathbf{B}$ matrices as discussed above ($\alpha_{gy} = \alpha_{\tau g} = \beta_{g\tau} = 0$), we then estimate $\beta_{\tau\tau}$ by which we scale the IRFs to a shock of ε^τ equal to 1% of GDP, thus excluding any initial feedback via u^y or u^g.

In terms of social security contributions, with reasonable signs for $\alpha_{\tau y}(> 0)$ and $\alpha_{y\tau}(< 0)$, the reported GDP reaction is stronger for (II) than for (I), as it requires an increase in statutory revenues of more than 1% of GDP to raise 1% of GDP in effective revenues. The same holds true for social security expenditures, where we expect $\alpha_{\tau y}(< 0)$ and $\alpha_{y\tau}(> 0)$. Increasing statutory expenditures by 1% of GDP is expected to increase effective spending by less than 1% of GDP. In both cases the measured impact multiplier will be higher for option (II) than for (I). Note that the difference between (I) and (II) 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 (I) for our baseline estimates, but also test the alternative choice (II). It turns out that the difference is small in our case due to the relatively low elasticities. In line with the theoretical channel, the impact multipliers are slightly upward-biased, when choosing the $\mathfrak{B}$ model.The GDP response for both sides of social security hardly differs from our baseline results. It is a bit higher on impact. This reflects the fact that the effective impact value of τ is fixed to 1% of GDP and thus the shock size is a little larger than in the baseline case.

Figure 5c presents the GDP responses for the MR approach using a more comprehensive VAR model where we include both social security expenditures and revenues at once. In this case, we have to decide on the within-quarter causal order between the two sides of the budget. We impose a Choleski ordering with expenditures ranked prior to revenues. The effects of shocks to social security expenditures remain quite robust, even though error bands are a little bit wider. The effects for revenues are somewhat lower than in the baseline and even die out more quickly. Note that reversing the order would not change the results much, the difference between revenue and spending multipliers would even be slightly increased (not shown).

The related literature (Romer and Romer 2016; Hayo and Uhl 2014; Cloyne 2013; Romer and Romer 2010) has relied on dynamic multiplier functions from VAR or ARDL models as described in equation (6) as opposed to the proxy SVAR approach. On the right hand side, w_t are reduced-form residuals, v again includes a constant, a reunification dummy and a financial crisis dummy like in section 3. $\boldsymbol{\lambda}(L)m_t$ is a 4th-order lag polynomial containing the (lagged) exogenous shock series and its coefficients. Figure 5d shows the dynamic multiplier effects of y to a shock in social security revenues and expenditures, respectively. The impact multipliers for social security revenues and expenditures hardly differ from our baseline results. Nonetheless, the effects become larger in the subsequent quarters in the dynamic multiplier framework. In the case of revenues, the increase in size is substantial and peaks at two, one year after the shock, before the effects again slowly decline towards zero. Note, however, the large error bands when using dynamic multipliers, which are basically in line with (Romer and Romer 2016; Hayo and Uhl 2014).

Turning to robustness checks regarding alternative specifications of the shock series and sample, Figure 6a first deals with the issue of fiscal foresight by restricting the shock series to changes where the period between publication and implementation date does not exceed 90 days, following MR. Using only non-anticipated shocks hardly affects our baseline results in the case of social security revenues. For changes to benefits, multiplier effects increase on impact to 1.5 and remain statistically significantly different to our baseline results for six quarters. Such an increase is plausible if for the anticipated changes we miss out spending increases of beneficiaries that expect higher future benefits. In terms of contributions there may be a countervailing effect that might explain why there is no measurable change. Consider an anticipated lowering of contributions: this may increase expected lifetime income and therefore lead to spending increases that we miss out; but it may on the other hand lead to a postponement of planned economic activity until the cut in contribution actually kicks in. We would then also miss out the contraction of economic activity prior to implementation.

Figure 6b shows GDP responses for estimations with the MR-approach for the full discretionary series (endogenous + exogenous). Given that endogenous discretionary reactions are counter-cyclical, one would expect the IRFs to be downward-biased as compared to the merely exogenous shocks. Estimating the responses for the full expenditure shock series yields the expected lower results, statistically significantly different to baseline for the first two quarters. For revenues, however, we find somewhat higher multipliers with the full shock series, but the difference to our baseline results is insignificant. This may reflect that contribution rates to the social security system have sometimes been changed in a procyclical manner in Germany.

Figures 6c and 6d show the GDP response, when shocks are censored to either below or above 0.3% of GDP. The GDP response to big revenue shocks is somewhat lower, only after 10 quarters the difference compared to small shocks turns insignificant. The impact multiplier for small expenditure shocks is comparatively large with 1.8, afterwards the effects again continuously decline towards zero. These differences may be due to small sample issues.

Finally, we split the sample into two sub-periods (Figures 6e and 6f), an early years sample spreading from 1974 to 1999 and second, the more recent period from 1991 to 2013.[5] Impact multipliers for revenues for both sub-periods are larger than one and thus higher as compared to baseline results. The difference of the point estimates between both sub-periods is statistically significant for the impact quarter and again one year after the exogenous shock, with effects in the earlier period dying out more quickly than in the later one. Turning to expenditures, the GDP response is significantly smaller and quickly approaches zero for early years of our sample, while it is somewhat higher on impact with regard to more recent years.

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

6 Conclusion

Following the narrative approach of identification of exogenous fiscal policy shocks (Romer and Romer 2010), we have constructed a rich narrative dataset for Germany by coding a shock series for social security contributions, benefits and transfers derived from official documents of major legislative changes in pensions, health care, long-term

[5]Note that at a split by half, the SVAR did not solve, most likely due to some small sample problems. We therefore chose reunification and introduction of the Euro as reasonable splitting dates.

care, unemployment insurance and basic social security. Based on quarterly data for 1974q1 to 2013q4 we have estimated the multiplier effects of changes to social security contributions and expenditures of this bottom-up identification within a proxy SVAR framework developed by Mertens and Ravn (2013). We compare them with estimates from a traditional cyclical-adjustment identification framework following Blanchard and Perotti (2002).

We find social security multipliers for Germany between 0.5 and 1.5 for shocks to both sides of the social security budget as well as different specifications of our model and shock series. Qualitatively on impact the effects for changes to benefits are in line with estimates by Romer and Romer (2016) based on US data, however, the effects for Germany are more persistent. Moreover, as opposed to Romer and Romer (2016), the effects are not only contained by private consumption expenditures but spread to output more generally. The GDP response to changes in social security contributions is a bit smaller on impact and peters out relatively quickly. The difference between the effects for contributions and benefits can be rationalized by heterogeneous marginal propensities to consume (MPC), with poorer households, who are likely net beneficiaries, having a higher MPC than richer households, who are more likely net contributors (Jappelli and Pistaferri 2014).

The multipliers as well as the implied elasticities of the social security components do square well with estimates from a Blanchard-Perotti (BP) SVAR identification. The huge discrepancy of tax multipliers using either the narrative or the cyclical adjustment approach that is discussed in MR and Caldara and Kamps (2017) does not carry over to our social security data.

In conclusion, we find that expansionary social security changes have a positive short-to-medium-term impact on GDP for Germany, somewhere in the middle of the range of multipliers in the literature (Gechert 2015). Given the size of the social security system and its frequent substantial law changes, they are likely as relevant for macroeconomic dynamics as are those to the tax system and general government spending, making them a relevant subject for future research.

References

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 (2008), What are the effects of fiscal policy shocks? A VAR-based comparative analysis, 877, European Central Bank working paper series.

Caldara, D. and C. Kamps (2017), The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers, *Review of Economic Studies 84*(3), 1015–1040.

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 (2018), A Narrative Account of Legislated Social Security Changes for Germany, Mimeo.

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

Gonçalvez, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics 123*, 89–120.

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

Jappelli, T. and L. Pistaferri (2014), Fiscal Policy and MPC Heterogeneity, *American Economic Journal: Macroeconomics 6*(4), 107–136.

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.

Price, R., T.-T. Dang and Y. Guillemette (2014), New Tax and Expenditure Elasticity Estimates for EU Budget Surveillance, 1174, 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 (2016), Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Increases, 1952-1991, *American Economic Journal: Macroeconomics 8*(4), 1–42.

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, [15/01/2014], http://www.portal-sozialpolitik.de.

Wilcox, D. (1989), Social Security Benefits, Consumption Expenditures, and the Life Cycle Hypothesis, *Journal of Political Economy 97*(2), 288–304.