Time variation in the dynamic effects of unanticipated changes in tax policy

Joris de Wind

Time variation in the dynamic effects of unanticipated changes in tax policy[*]

Joris de Wind[†]

March, 2014

Abstract

Using a structural vector autoregression with time-varying parameters, I analyze to what extent the dynamic effects of unanticipated changes in tax policy have changed structurally over the post World War II period in the United States. The estimated time variation points to a permanent decline in the tax multiplier as well as a faster response of the economy. Despite the permanent decline, the estimated tax multiplier is still at the higher end of the range of existing empirical estimates, which is consistent with Mertens and Ravn (2013b), whose identification strategy I follow. Furthermore, the estimated time variation also suggests that fiscal policy has become more countercyclical over time. In particular, spending policy used to be procyclical and has become countercyclical after the beginning of the 1990s, whereas tax policy already used to be countercyclical and has become even more countercyclical over time.

Nederlandse samenvatting (summary in Dutch)

In dit paper onderzoek ik in welke mate de dynamische effecten van onverwachte veranderingen in belastingbeleid structureel zijn veranderd gedurende de periode na de Tweede Wereldoorlog in de Verenigde Staten. Het gebruikte empirische raamwerk is gebaseerd op De Wind en Gambetti (2014), waarin een nieuwe model opzet voor vector autoregressieve (VAR) modellen met tijdsvariërende parameters is ontwikkeld. Voor de identificatie van de structurele belastingschok heb ik de identificatie strategie van Mertens en Ravn (2013b) gevolgd, waarbij gebruik wordt gemaakt van informatie over onverwachte veranderingen in belastingbeleid (via een zogenaamde narratieve tijdreeks die geconstrueerd is op basis van veranderingen in de wetgeving). Het kenmerkende onderscheid met hun paper is dat ik gebruik heb gemaakt van een tijdsvariërende parameter VAR model, terwijl zij daarentegen gebruik hebben gemaakt van een VAR model met constante parameters.

De resultaten duiden op een permanente afname van de Amerikaanse belastingmultiplier alsook een snellere reactie van de economie. Ondanks de permanente afname bevindt de geschatte belastingmultiplier zich nog steeds aan de bovenkant van de range van bestaande empirische schattingen voor de Verenigde Staten, wat consistent is met Mertens en Ravn (2013b).

Daarnaast heb ik ook geanalyseerd in hoeverre begrotingsbeleid anti-cyclischer is geworden in de loop der tijd. De resultaten geven aan dat het bestedingsbeleid pro-cyclisch was en anti-cyclisch is geworden na het begin van de jaren 1990, terwijl het belastingbeleid reeds anti-cyclisch was en in de loop der tijd zelfs nog anti-cyclischer is geworden.

1   Introduction

Fiscal stimulus has been rediscovered as an important instrument to stimulate the economy. After the collapse of Lehman Brothers and the following sharp drops in real activity, policy makers in several countries have responded with fiscal stimulus packages to counteract the slowdown of the economy. The reaction of the United States was particularly vigorous and has resulted in the Economic Stimulus Act of 2008 and the American Recovery and Reinvestment Act of 2009. Further efforts by President Obama such as the American Jobs Act were blocked in 2011 by the Republicans, who had acquired the majority in the House of Representatives in 2010.

The strong countercyclical response to the 2007–2008 financial crisis, and the general consensus in support of the large fiscal stimulus, is a marked change from earlier episodes. Auerbach (2009) argues that the change towards countercyclical fiscal policy predates the 2008–2009 stimulus packages, as the response to the early 2000s recession included countercyclical bonus depreciation investment incentives as opposed to the clearly procyclical responses to the early 1980s and early 1990s recessions.[1] Auerbach (2009) also presents evidence that fiscal policy has overall been countercyclical, although it is not clear how much of this result is driven by the exclusion of the early 1980s recession (and the subsequent procyclical fiscal response) from the estimation sample. In another paper, Auerbach (2008) identifies different budget control regimes in the United States and shows that fiscal policy has become more countercyclical after the introduction of the Budget Enforcement Act of 1990, based on (insignificant) sample-split results. In this paper, I present systematic evidence on the cyclicality of fiscal policy using a vector autoregression (VAR) with time-varying parameters, and indeed fiscal policy seems to have become more countercyclical over time.

In view of the increased use of countercyclical fiscal policy, it is important to understand the dynamic effects of changes in fiscal policy. How large are the effects on impact? After how many quarters do the effects reach their peak and how large is the peak? How fast do the effects die out thereafter? The empirical literature does not answer these questions with one voice. There is no consensus regarding the dynamic effects of changes in government expenditure nor regarding the dynamic effects of changes in tax policy. In this paper, I focus on the dynamic effects of unanticipated changes in tax policy.

Most studies agree on the direction of the effects of tax increases, yet there is not much consensus on the size of the effects. Most studies belief that tax increases are generally contractionary, yet estimates

---

[1]Countercyclical fiscal policy goes against the business cycle (stimulus in downturn, austerity in upswing), whereas procyclical fiscal policy goes with the business cycle (austerity in downturn, stimulus in upswing).

for the US tax multiplier range from close to zero to more than four.[2] The broad range reflects numerous differences in methodology, identification assumptions, model specification, and sample coverage. In the empirical literature, there are two main methodological approaches to the estimation of the US tax multiplier. In the structural VAR approach, tax shocks are identified by imposing short-run restrictions as in the seminal contribution by Blanchard and Perotti (2002) or by imposing sign restrictions as in Mountford and Uhlig (2009). In the narrative approach, a narrative series of observable (legislated) changes in tax policy, such as the one constructed by Romer and Romer (2009), is used to measure tax shocks directly. Recently, the two approaches are combined by Mertens and Ravn (2013a) as well as Mertens and Ravn (2013b), who use a narrative series as a proxy that imperfectly measures tax shocks in a structural VAR model. In particular, they assume that the narrative series is correlated with the structural tax shock but uncorrelated with the other structural shocks, which is sufficient to identify the structural tax shock and to back out the tax multiplier. Mertens and Ravn (2013b) also give an overview of existing empirical estimates for the US tax multiplier, and with their "proxy structural VAR model," nesting both the structural VAR approach as well as the narrative approach, they can reconcile their own estimate for the US tax multiplier with the pre-existing ones.

In this paper, I have analyzed what the effects are of unanticipated tax changes and, in particular, to what extent these effects have *changed* structurally over the post World War II period in the United States. The main results of this paper are: (i) there has been a permanent decline in the tax multiplier, from more than four at the beginning of the sample period to around three at the end of the sample period and (ii) the response of the economy has become much faster, reaching its peak response after six quarters at the beginning of the sample period as opposed to four quarters at the end of the sample period. Despite the permanent decline, the estimated tax multiplier is still at the higher end of the range of existing empirical estimates, which is consistent with Mertens and Ravn (2013b), whose identification strategy I have followed.

The permanent decline in the tax multiplier and faster response of the economy found in this paper are consistent with the sample-split results of Mertens and Ravn (2013b), who have applied their proxy structural VAR methodology to compare the dynamic effects of an unanticipated tax shock in the post-1980 period with the dynamic effects in the pre-1980 period. Mertens and Ravn (2013b) use the subset

---

[2]The tax multiplier is defined as the ratio between the change in GDP (in dollars) and the autonomous change in tax revenues (in dollars) that triggered it. The tax multiplier is a dynamic concept and wherever I am not specific about the horizon of the tax multiplier, I refer to the tax multiplier at its peak (as opposed to the tax multiplier on impact).

of the Romer and Romer (2009) narrative series that includes only unanticipated tax changes. This still leaves 26 nonzero elements for the full sample, yet only nine nonzero elements for the second sub-sample that starts in 1980. The identification procedure, described above, exploits the informational content of the nonzero elements in the narrative series of unanticipated tax changes. Not surprisingly, the sample-split results of Mertens and Ravn (2013b) are not robust with such a low number of nonzero elements, as shown in this paper.

As an alternative for their sample-split VAR, I have extended the analysis of Mertens and Ravn (2013b) by allowing for time variation in the reduced-form parameters of the VAR model. The reduced-form model specification is, apart from the time-varying parameters, unchanged with the same set of variables and number of lags. Also the proxy identification procedure is the same, ditto the narrative series of unanticipated tax changes. For an extensive discussion of the benefits of time-varying VARs, see De Wind and Gambetti (2014) and the references therein. Here, I only would like to emphasize that time-varying VARs are more appropriate for analyzing time variation than sample-split VARs. In particular, with time-varying VARs the timing and number of sub-samples are endogenously determined by the data, whereas with sample-split VARs this is simply imposed by the researcher without proper empirical support. Moreover, the sample-split method is particularly sensitive in combination with the proxy identification procedure due to the low number of nonzero elements in the narrative series. Despite the differences in methodology, the results from the time-varying VAR model used in this paper are consistent with the sample-split results of Mertens and Ravn (2013b).

Finally, my analysis focuses on the structural change in the effects of discretionary fiscal policy (in particular, unanticipated tax changes), which is complementary to the literature that argues that the effects of discretionary fiscal policy depend on the state of the business cycle. There is both empirical as well as theoretical work arguing that fiscal multipliers are larger when the economy is in recession. An empirical example is Auerbach and Gorodnichenko (2012), who use a smooth transition structural VAR model and estimate that fiscal multipliers are larger in the recessionary regime. A theoretical example is Christiano, Eichenbaum, and Rebelo (2011), who develop a DSGE model and find that fiscal multipliers are larger at the zero lower bound on nominal interest rates, which happens only in recessionary times.

The organization of the rest of this paper is as follows. In section 2, I replicate the proxy structural VAR results of Mertens and Ravn (2013b) and show that their sample-split results are not robust. In section 3, I describe the time-varying VAR model used in this paper and give the priors used in the Bayesian estimation. In section 4, I first present evidence that fiscal policy has become more countercyclical over

time and then I present evidence on the permanent decline in the tax multiplier and faster response of the economy. I also give various robustness checks. Finally, section 5 concludes.

2 Mertens and Ravn (2013b) replication

In this section, I replicate the proxy structural VAR results of Mertens and Ravn (2013b). In particular, I focus on the dynamic effects of unanticipated tax changes and show that the sample-split results of Mertens and Ravn (2013b) are not robust along this dimension. Before presenting the results, I first describe the reduced-form VAR model specification and the proxy identification procedure.

2.1 Reduced-form VAR model specification and data description

Mertens and Ravn (2013b) use the same reduced-form VAR model specification as Blanchard and Perotti (2002), which is the benchmark specification in the literature. The model contains quarterly US data on tax revenues T_t, government spending G_t, and output Y_t. It is important to model tax revenues and government spending jointly because they are presumably not independent. Moreover, it is important to include four lags in the model since some taxes, such as those on corporate income, are often paid with substantial delays and, in addition, these delays are quarter dependent, an argument put forward by Blanchard and Perotti (2002).

The variables in the trivariate VAR are specified in levels. In particular, the vector of endogenous variables $y_t = [T_t, G_t, Y_t]'$ is modeled by

$$y_t = \alpha' d_t + \sum_{i=1}^{p} B_i y_{t-i} + u_t \tag{1}$$

where $\{B_i\}_{i=1}^{p}$ are $n \times n$ matrices of autoregressive parameters in the model with $n = 3$ variables and $p = 4$ lags. The $n \times 1$ vector u_t with reduced-form shocks is assumed to be distributed according to the normal distribution $u_t \sim N(0, \Sigma)$. To account for low-frequency properties of the data, the model includes a linear and quadratic time trend in the vector d_t with deterministic components, in addition to a constant and a dummy for 1975Q2. The coefficients for the deterministic components are represented by the matrix α of suitable dimensions.[3]

---

[3]Blanchard and Perotti (2002) also include four lags of the 1975Q2 dummy, which are left out by Mertens and Ravn (2013b).

The reduced-form VAR model is estimated on quarterly US data from the Bureau of Economic Analysis. In particular, the data for federal tax revenues, federal government consumption and investment expenditures, and GDP are taken from the national income and product account tables. All variables are in log real per capita terms and the sample runs from 1950Q1 up to and including 2006Q4. For more details about the data, see Mertens and Ravn (2013b).

Mertens and Ravn (2013b) use a narrative series of observable, legislated changes in tax policy for the identification of unanticipated tax shocks. In particular, they use the narrative series constructed by Romer and Romer (2009) regarding legislated changes in tax policy on the federal level,[4] yet only include the subset with changes that are unanticipated. Mertens and Ravn (2013b) make the assumption that changes in tax policy are unanticipated when the legislation and implementation date are less than one quarter apart, leaving 26 nonzero elements in their proxy. Figure 1 plots the Mertens and Ravn (2013b) proxy, which is scaled by previous quarter nominal GDP, together with the revenues-GDP ratio.

2.2 Proxy identification procedure

Mertens and Ravn (2013b) use the proxy identification procedure they had developed earlier in Mertens and Ravn (2013a). Although the proxy identification procedure actually originates to slides by Stock and Watson (2008) and was laid out at approximately the same time in Stock and Watson (2012), it was developed independently by Mertens and Ravn (2013a).

The key idea of the proxy identification procedure is to use narrative series as proxies that imperfectly measure the structural shocks in a structural VAR model. The structural shocks can be identified by imposing moment conditions between the narrative series and the structural shocks. Although the proxy identification procedure is more general, it is explained here for the case with one proxy for the identification of one structural shock (which, in the current application, is labeled as structural tax shock).[5]

As is standard in the structural VAR literature, the identification procedure starts with the assumption that the $n \times 1$ vector u_t with reduced-form shocks depends linearly on the $n \times 1$ vector ε_t with structural

---

[4]Romer and Romer (2009) have selected legislated changes in tax policy that are motivated either as ideological or as arising from inherited debt concerns, so that the changes can be considered as exogenous to the fluctuations in the economy.

[5]For the case with multiple proxies for the identification of one structural shock, see Stock and Watson (2012). For the case with k proxies for the identification of k structural shocks, see Mertens and Ravn (2013a).

shocks so that

$$u_t = B\varepsilon_t \tag{2}$$

where B is an $n \times n$ impact matrix. The structural shocks are, by nature, orthogonal to each other and the structural shocks can, without loss of generality, be normalized to have unit variance, implying that

$$\Sigma = BB' \tag{3}$$

which amounts to $\frac{n(n+1)}{2}$ restrictions on the impact matrix B. This is, however, not sufficient to fully identify the impact matrix, which has n^2 free elements, and therefore additional identifying restrictions are needed. Typically, tax shocks are identified by imposing short-run restrictions as in the seminal contribution by Blanchard and Perotti (2002) or by imposing sign restrictions as in Mountford and Uhlig (2009).

The proxy identification procedure instead exploits the informational content of a narrative series to obtain the needed additional identifying restrictions. In particular, it is assumed that the narrative series is correlated with the structural tax shock but uncorrelated with the other structural shocks, thus

$$\mathrm{E}\left(\varepsilon_t^T m_t\right) = \phi \neq 0 \tag{4a}$$

$$\mathrm{E}\left(\varepsilon_t^O m_t\right) = 0 \tag{4b}$$

where the proxy is denoted by the scalar m_t, the structural tax shock by ε_t^T, and the vector with the other structural shocks by ε_t^O. The structural shocks are ordered as $\varepsilon_t = [\varepsilon_t^T, (\varepsilon_t^O)']'$. Furthermore, the proxy is assumed to have a zero mean, which can without loss of generality be satisfied by subtracting the mean of the nonzero elements from the nonzero elements.[6]

The aforementioned moment conditions are sufficient to identify the structural tax shock and the associated impulse response function. The derivation of the impulse response function and the procedure to back out the full history of the structural tax shock are discussed in turn.

First, note that for the identification of the impulse response function of the structural tax shock we only need the first column of the impact matrix B, so that it is useful to partition $B = [\beta_T, \beta_O]$ where

---

[6]The proxy identification procedure also requires that the proxy is uncorrelated with the history of y_t. Nevertheless, note that it is always possible to use the projection error of the projection of the proxy on the history of y_t, which is orthogonal by construction.

β_T represents the first column and β_O represents the other columns. Using the aforementioned moment conditions we can derive

$$\mathrm{E}\left(u_t m_t\right) = B\,\mathrm{E}\left(\varepsilon_t m_t\right) = \beta_T\,\mathrm{E}\left(\varepsilon_t^T m_t\right) + \beta_O\,\mathrm{E}\left(\varepsilon_t^O m_t\right) = \beta_T \phi \tag{5}$$

which implies that the covariance between the reduced-form shocks and the narrative series, for which we can simply get a sample estimate, pins down the impact vector β_T up to scaling. The covariance parameter ϕ, which serves in the above equation as constant-of-proportionality, is still to be estimated, although we do not need it for all purposes. In particular, the shape of the impulse response function, and with that the size of the tax multiplier, does not depend on the scaling. Regarding the impulse response function, we only need the constant-of-proportionality if we want to show the response to a (negative) one-standard deviation structural tax shock, but not if the magnitude of the impulse is scaled otherwise.

In the current application, the impulse to the structural tax shock is scaled such that it amounts to a decrease in tax revenues of one percent of GDP. More specifically, the impulse to the structural tax shock is first scaled such that it amounts to a one-percent decrease in tax revenues and then it is divided by the sample average of the ratio between tax revenues and GDP, which is approximately 17.5%. This scaling gives the response of GDP the interpretation of tax multiplier, i.e. the ratio between the change in GDP (in dollars) and the autonomous change in tax revenues (in dollars) that triggered it. The tax multiplier is a dynamic concept and the full dynamic pattern can be read directly from the impulse response functions reported in this paper. Wherever I am not specific about the horizon of the tax multiplier, I refer to the tax multiplier at its peak (as opposed to the tax multiplier on impact).

Altogether, the impulse response function of the structural tax shock can be estimated using the following steps:

1. Estimate the reduced-form VAR model by ordinary least squares

2. Estimate the vector with covariances between the reduced-form shocks and the narrative series, which pins down the first column of the impact matrix up to scaling

3. Scale the first column of the impact matrix such that the first element is equal to the negative of the reciprocal of the sample average of the ratio between tax revenues and GDP, which specifies the impulse to the economy

4. Enter the impulse to the economy as a one-time reduced-form shock and iterate on the reduced-form VAR model to determine the impulse response function

Second, to back out the full history of the structural tax shock we need the constant-of-proportionality, which was not needed for the derivation of the impulse response function. For the case discussed here, with one proxy for the identification of one structural shock, the constant-of-proportionality is available without additional identification assumptions. Start the derivation by partitioning the impact matrix as

$$B = \begin{pmatrix} \beta_{TT} & \beta_{TO} \\ \beta_{OT} & \beta_{OO} \end{pmatrix} \tag{6}$$

where, in the case discussed here, the upper left block is a scalar, the lower left block is a column vector, the upper right block is a row vector, and the lower right block is a square matrix. The covariance matrix of the reduced-form shocks is partitioned analogously, where in addition $\Sigma_{TO} = \Sigma'_{OT}$. Mertens and Ravn (2013a) provide an expression for the upper left block, which in the scalar case discussed here can be simplified to

$$\beta_{TT} = \sqrt{\Sigma_{TT} - \left(\Sigma_{OT} - \beta_{OT}\beta_{TT}^{-1}\Sigma_{TT}\right)' \Gamma^{-1} \left(\Sigma_{OT} - \beta_{OT}\beta_{TT}^{-1}\Sigma_{TT}\right)} \tag{7}$$

where $\Gamma = \beta_{OT}\beta_{TT}^{-1}\Sigma_{TT} \left(\beta_{OT}\beta_{TT}^{-1}\right)' - \Sigma_{OT} \left(\beta_{OT}\beta_{TT}^{-1}\right)' - \beta_{OT}\beta_{TT}^{-1}\Sigma'_{OT} + \Sigma_{OO}$.[7] The essential ingredients for this expression are the various blocks of Σ and the term $\beta_{OT}\beta_{TT}^{-1}$. The latter term is pinned down by β_T, the first column of the impact matrix, even though we do not yet know the constant-of-proportionality. In appendix A, I provide a derivation for the expression of Mertens and Ravn (2013a).

Now that we have an expression for the scalar in the upper left block, we can back out the constant-of-proportionality and in turn the rest of the first column of the impact matrix. To back out the full history of the structural tax shock, express the structural shocks in terms of the reduced-form shocks

$$\varepsilon_t = B^{-1}u_t \tag{8}$$

which reveals that we need the first row of the inverse impact matrix. In appendix B, I show how to derive the first row of the inverse impact matrix without making additional identification assumptions.

Finally, the proxy identification procedure seems to be an appealing methodology as it integrates the

---

[7]Mertens and Ravn (2013a) provide the following general expression for the upper left block

$$\beta_{TT}\beta'_{TT} = \Sigma_{TT} - \left(\Sigma_{OT} - \beta_{OT}\beta_{TT}^{-1}\Sigma_{TT}\right)' \Gamma^{-1} \left(\Sigma_{OT} - \beta_{OT}\beta_{TT}^{-1}\Sigma_{TT}\right)$$

which can only be simplified to the expression in the main text in the case that β_{TT} is a scalar and otherwise additional identification assumptions are needed to back out β_{TT}.

narrative approach in the standard structural VAR framework and combines the attractive features of
the two approaches. Relative to the narrative approach, it allows for measurement error in the narrative
series which seems like a good idea as it is unlikely that a narrative series can be constructed perfectly
and without any judgement. Relative to the structural VAR approach, the informational content of a
narrative series is exploited rather than having to make identification assumptions on which there is no
consensus.

2.3   Full-sample results

I have applied the proxy structural VAR methodology to the full sample and I can successfully replicate
the proxy structural VAR results of Mertens and Ravn (2013b). In my reporting, I focus on the dynamic
effects of unanticipated changes in tax policy, as summarized by the impulse response function of the
structural tax shock.

Figure 2 plots the impulse response function of the structural tax shock and is basically identical to
the figure presented by Mertens and Ravn (2013b). The impulse to the structural tax shock is scaled
such that it amounts to a 5.7% decrease in tax revenues, which is on average one percent of GDP. In
response to the structural tax shock, GDP increases on impact by 2.0% and reaches its peak after five
quarters with a 3.2% increase over its trend. As explained in the previous subsection, these numbers can
be interpreted as tax multipliers and lie at the higher end of the range of existing estimates for the US
tax multiplier. For a more extensive comparison, see Mertens and Ravn (2013b) who reconcile their own
estimate for the US tax multiplier with the pre-existing ones.

2.4   Sample-split results

I have applied the proxy structural VAR methodology to the pre-1980 period and the post-1980 period
and I can successfully replicate the sample-split results of Mertens and Ravn (2013b). As for the full
sample, I only report on the impulse response function of the structural tax shock.[8]

Figure 3 plots the impulse response function of the structural tax shock for the pre-1980 period and
the post-1980 period and is basically identical to the figure presented by Mertens and Ravn (2013b). For
both periods, the impulse to the structural tax shock is scaled such that it amounts to a decrease in tax
revenues of one percent of GDP. This amounts to a decrease in tax revenues of 5.8% and 5.6% for the

---

[8]The pre-1980 sample runs from 1950Q1 up to and including 1979Q4. The post-1980 sample runs from 1980Q1 up to
and including 2006Q4.

pre-1980 and post-1980 period, respectively. Although the impact on GDP has not changed much, the response of GDP reaches its peak much quicker in the post-1980 period than in the pre-1980 period and the peak is also much lower (2.0% increase over trend after four quarters as opposed to 2.7% increase over trend after six quarters). Moreover, the reversion to the trend is also much quicker in the post-1980 period than in the pre-1980 period. This suggests that there is time variation in the dynamic effects of unanticipated changes in tax policy (which is why I extend the analysis of Mertens and Ravn (2013b) by allowing for time variation in the model parameters, to which the remainder of this paper is dedicated).

Nevertheless, it turns out that the evidence for time variation put forward by Mertens and Ravn (2013b) is not robust. In particular, their sample-split figure, which is replicated here as figure 3, is very sensitive to the timing of the split of the sample. For example, if the early sub-sample had lasted one more year, the impulse response function of the structural tax shock would have looked rather different, which is demonstrated in figure 4. With the extra year, the early sub-sample would have included the nonzero proxy element for 1980Q2, which is a high-leverage point (only three elements are further away from the mean and many elements are much closer to the mean) that turns out to be influential. The sensitivity is not really surprising given the low number of nonzero elements in the proxy, with just sixteen of them for the pre-1980 period. Clearly, the sample-split results of Mertens and Ravn (2013b) are not robust, although the question whether or not there is time variation can still be answered in the affirmative.

2.5 Alternative strategy for the sample split

In this subsection, I propose an alternative strategy for the sample split that generates sample-split results that are robust, in contrast to the sample-split results of Mertens and Ravn (2013b). The key idea is to use the sample split for the estimation of the reduced-form VAR model but not for the identification of the structural tax shock. In particular, the estimation of the covariance between the reduced-form shocks and the narrative series, which is the main ingredient in the proxy identification procedure, should be based on the two sub-samples stacked together, so as to avoid that this covariance is estimated on the basis of too few nonzero proxy elements (which is, in fact, the Achilles' heel of the Mertens and Ravn (2013b) strategy). Furthermore, the scaling part of the proxy identification procedure can again be dealt with separately for the two sub-samples.

Figure 5 plots the impulse response function of the structural tax shock for the pre-1980 period and the post-1980 period based on my alternative sample-split strategy. Despite the sensitivity of the Mertens and Ravn (2013b) sample-split figure, the alternative impulse response function actually turns out to be

very similar. The impact of the structural tax shock on GDP has again not changed much over time, while the dynamic patterns after the impact actually depend only on the reduced-form VAR models, which are, by construction, unaffected relative to the Mertens and Ravn (2013b) sample-split strategy.

3 Time-varying VAR model specifications

In this section, I give the two time-varying VAR model specifications used in this paper. The first one does not include stochastic volatility, whereas the second one does. The two model specifications only differ to the extent that the covariance matrix of the reduced-form shocks is time-invariant in the case without stochastic volatility and time-varying in the case with stochastic volatility. Henceforth, the two model specifications are referred to as TV-VAR-NO and TV-VAR-SV, respectively. Before presenting the two model specifications (in the next two subsections), I have three general remarks.

First, the identification of the impulse response function of the structural tax shock, which is based on the proxy identification procedure as described in section 2.2, does not depend on the covariance matrix of the reduced-form shocks.[9] Therefore, allowing for time variation in the covariance matrix of the reduced-form shocks would not contribute directly to the time variation in the impulse response function of the structural tax shock. The main results of this paper are, for this reason, based on the TV-VAR-NO specification. Nevertheless, I also present results based on the TV-VAR-SV specification, which is an important robustness check as the incorporation of stochastic volatility could possibly matter in an indirect manner (as explained in section 4.3.1).

Second, the covariance matrix of the innovations to the time-varying parameters is, in both specifications, assumed to be of reduced rank. In particular, I use the reduced-rank model setup as developed in De Wind and Gambetti (2014). The reduced-rank model setup is more parsimonious than the standard (full-rank) model setup, and this is particularly advantageous for analyzing larger models including more variables and/or more lags. In the current application, I include four lags in the model which is standard in the fiscal VAR literature, yet rather large for time-varying VARs. For an extensive discussion of time-varying VARs, the benefits of the rank reduction as well as the Bayesian estimation procedures, see De Wind and Gambetti (2014). Besides, apart from the rank reduction, the model setup is basically identical to the time-varying VAR model setup of Primiceri (2005) except for one minor difference as

---

[9]In fact, the reduced-form shocks themselves enter the proxy identification procedure and their covariance matrix is only needed to back out the constant-of-proportionality (which is not needed to identify the shape of the impulse response function).

explained in footnote 10.

Third, the deterministic components (linear and quadratic time trends as well as a dummy for 1975Q2) that are included in the VAR model specification of Mertens and Ravn (2013b) are excluded from the time-varying VAR model specifications used in this paper. Instead I have pre-filtered the data, implemented by a linear regression of the data on a constant and the time trends. I also present results based on other trend specifications and the results are robust along this dimension.

3.1 TV-VAR-NO

Consider the model with n variables and p lags

$$y_t = c_t + \sum_{i=1}^{p} B_{i,t} y_{t-i} + u_t \tag{9}$$

where y_t is an $n \times 1$ vector of observed endogenous variables, c_t is an $n \times 1$ vector of time-varying constants, $\{B_{i,t}\}_{i=1}^{p}$ are $n \times n$ matrices of time-varying autoregressive parameters, and u_t is an $n \times 1$ vector of shocks. Shocks are assumed to be distributed according to the normal distribution $u_t \sim N(0, \Sigma)$.

Let $\theta_t \equiv \text{vec}([c_t, \{B_{i,t}\}_{i=1}^{p}]')$ denote the vector of time-varying parameters, where vec is the column stacking operator. The dimension of θ_t is k by 1 with $k = n + pn^2$. The law of motion for θ_t is assumed to be equal to

$$\theta_t = \theta_{t-1} + \nu_{\theta,t} \tag{10}$$

where $\nu_{\theta,t}$ is a $k \times 1$ vector of shocks. Shocks are assumed to be distributed according to the normal distribution $\nu_{\theta,t} \sim N(0, Q_\theta)$, where the covariance matrix Q_θ is assumed to be of reduced rank, i.e. the rank is chosen to be equal to $\text{rank}(Q_\theta) = q_\theta \leq k$. The rank reduction implies cross-equation restrictions that amount to a reduction in the number of underlying factors driving the time-varying parameters. In particular, the $q_\theta \times 1$ vector of underlying factors $\widetilde{\theta}_t$ can, without loss of generality, be defined implicitly by

$$\theta_t \equiv \Lambda_\theta \widetilde{\theta}_t + M_\theta \theta_0 \tag{11}$$

where Λ_θ is a $k \times q_\theta$ matrix of factor loadings and $M_\theta = I_k - \Lambda_\theta (\Lambda_\theta' \Lambda_\theta)^{-1} \Lambda_\theta'$ is the projection matrix onto the left null space of Λ_θ. The law of motion for the underlying factors is given by

$$\widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \upsilon_{\theta,t} \tag{12}$$

where $\upsilon_{\theta,t}$ is a $q_\theta \times 1$ vector of standard normally distributed shocks, which is assumed to be uncorrelated with the reduced-form shocks in the VAR equation. For details on the derivation of the underlying factor structure as well as intuition on the decompositions, see De Wind and Gambetti (2014).

Next, define the $n \times k$ matrix of regressors $X'_t \equiv I_n \otimes [1, \{y'_{t-i}\}_{i=1}^p]$, where $\otimes$ denotes the Kronecker product, and rewrite the VAR equation in concise matrix form. Together with the law of motion for the underlying factors and the implicit definition of the underlying factors, the model specification can now be represented by the following state-space representation

$$y_t = X'_t \theta_t + u_t \tag{13a}$$

$$\widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \upsilon_{\theta,t} \qquad \theta_t \equiv \Lambda_\theta \widetilde{\theta}_t + M_\theta \theta_0 \tag{13b}$$

Finally, the Bayesian estimation procedure is outlined in De Wind and Gambetti (2014). For the priors and number of underlying factors used in the current application, see section 3.3.

3.2 TV-VAR-SV

The model specification with stochastic volatility is similar to the model specification without stochastic volatility, aside from the covariance matrix of the reduced-form shocks in the VAR equation which is now time-varying Σ_t rather than time-invariant Σ. Introducing stochastic volatility requires restrictions to make sure that the covariance matrix is always positive definite.

Cholesky decompose the covariance matrix $\Sigma_t = \left(A_t^{-1}\Omega_t\right)\left(A_t^{-1}\Omega_t\right)'$ where A_t^{-1} is a lower triangular matrix with ones on the main diagonal and Ω_t is a diagonal matrix. Let α_t be the vector of elements below the main diagonal of the matrix A_t stacked by rows. The dimension of α_t is r by 1 with $r = \frac{n(n-1)}{2}$. The law of motion for α_t is assumed to be equal to

$$\alpha_t = \alpha_{t-1} + \nu_{\alpha,t} \tag{14}$$

where $\nu_{\alpha,t}$ is an $r \times 1$ vector of shocks. Shocks are assumed to be distributed according to the normal distribution $\nu_{\alpha,t} \sim N(0, Q_\alpha)$.[10] Let σ_t denote the vector of diagonal elements of the matrix Ω_t. The

---

[10]Here, Q_α is a dense matrix, in contrast to Primiceri (2005). The reason why I have relaxed the block-diagonal assumption of Primiceri (2005) is explained in De Wind and Gambetti (2014).

dimension of σ_t is n by 1. The law of motion for σ_t is assumed to be equal to

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \nu_{\sigma,t} \tag{15}$$

where $\nu_{\sigma,t}$ is an $n \times 1$ vector of shocks. Shocks are assumed to be distributed according to the normal distribution $\nu_{\sigma,t} \sim N(0, Q_\sigma)$. As opposed to the time-varying parameters, I do not impose an underlying factor structure (reduce the rank of the covariance matrices) in the stochastic volatility part of the model setup. Note that the dimensions of the state vectors α_t and $\log(\sigma_t)$ are much smaller than the dimension of the state vector θ_t so that the benefits would have been much smaller.

Next, use the above Cholesky decomposition in order to decompose the shocks in the VAR equation as $u_t = A_t^{-1}\Omega_t\varepsilon_t$, where ε_t is an $n \times 1$ vector of standard normally distributed shocks, and rewrite the time-varying VAR in concise matrix form. Together with the laws of motion for the time-varying parameters (in terms of the underlying factor structure) and stochastic volatility, the model specification can now be represented by the following state-space representation

$$
\begin{align}
y_t &= X_t'\theta_t + A_t^{-1}\Omega_t\varepsilon_t \tag{16a}\\
\widetilde{\theta}_t &= \widetilde{\theta}_{t-1} + v_{\theta,t} \qquad\qquad \theta_t \equiv \Lambda_\theta\widetilde{\theta}_t + M_\theta\theta_0 \tag{16b}\\
\alpha_t &= \alpha_{t-1} + \nu_{\alpha,t} \tag{16c}\\
\log(\sigma_t) &= \log(\sigma_{t-1}) + \nu_{\sigma,t} \tag{16d}
\end{align}
$$

where, besides, the four vectors of shocks are assumed to be mutually uncorrelated.

Finally, the Bayesian estimation procedure is outlined in De Wind and Gambetti (2014). For the priors and number of underlying factors used in the current application, see section 3.3.

3.3   Prior settings

I have allocated a period of twelve years (from 1951Q3 up to and including 1963Q2) to calibrate the priors.[11] In particular, the training sample is used to estimate a time-invariant VAR from which the priors are constructed, which is the standard approach in the time-varying VAR literature. There is, however, an important improvement in the sense that I have used a Bayesian regression to estimate the

---

[11]The training sample, which consists of only twelve years of data, is somewhat sensitive to the data from 1950Q1 up to and including 1951Q2. Therefore, I have excluded these first six data points from the Mertens and Ravn (2013b) sample.

time-invariant VAR as opposed to the standard classical regression.[12] This is particularly important for the current application given the more than usual number of lags included in the model.[13] Apart from the Bayesian regression, the strategy for calibrating the priors is standard. The various prior components are discussed in turn, first for the main specification without stochastic volatility (TV-VAR-NO) and then for the alternative specification with stochastic volatility (TV-VAR-SV).

3.3.1 Prior settings TV-VAR-NO

The marginal prior on the covariance matrix of the reduced-form shocks in the VAR equation (Σ) is assumed to be an inverse-Wishart distribution with $n \times n$ scale matrix $\bar{R}$ and $n + 1$ degrees of freedom. The scale matrix $\bar{R}$ is chosen to be equal to $n+1$ times the time-invariant point estimate of the covariance matrix of the reduced-form shocks in the VAR equation in the training sample. Note that the multiplication with the degrees of freedom is because the scale matrix has a sum-of-squares interpretation in the inverse-Wishart distribution.

Next, the marginal prior on the starting values of the time-varying parameters (θ_0) is assumed to be a normal distribution with mean $\bar{\theta}$ and covariance matrix $\bar{P}_\theta$. The mean $\bar{\theta}$ is chosen to be equal to the time-invariant point estimate of the vector of VAR parameters in the training sample and the covariance matrix $\bar{P}_\theta$ is chosen to be equal to four times the covariance matrix of the time-invariant estimate of the vector of VAR parameters in the training sample.

Finally, the marginal prior on the covariance matrix of the shocks driving the time-varying parameters (Q_θ) needs a bit more explanation because of the rank reduction. It is my intention to follow the time-varying VAR literature as much as possible. Therefore, I start off from a prior on Q_θ^{full} for the full-rank model version and based thereon I construct the prior on Q_θ^{red} for the reduced-rank model version to imply the same amount of time variation in the parameters, just like in De Wind and Gambetti (2014).

---

[12]Regarding the Bayesian VAR, I have used a natural conjugate version of the so-called Minnesota prior, see e.g. Koop and Korobilis (2010) for an excellent survey article on Bayesian VARs. I have set the hyperparameter on the overall tightness to 0.2, as recommended by Sims and Zha (1998). Furthermore, I have centered the prior around univariate AR(1) processes with autocorrelation parameters of 0.9 as opposed to the more standard univariate random walks, which makes sense given that the data is pre-filtered.

[13]The Bayesian regression approach is more attractive than using a (much) longer training sample, as the training sample cannot be recycled in the main estimation of the time-varying VAR. Nevertheless, a much longer training sample is used by, for example, Baumeister and Peersman (2013). In their model with four variables and four lags, they use a training sample that is more than twice as long and throw away valuable data from the main estimation (which is particularly undesirable given their rather large model).

For the full-rank model version, the prior on Q_θ^{full} is standard. In particular, the prior is assumed to be an inverse-Wishart distribution with $k \times k$ scale matrix $\bar{Q}_\theta^{\text{full}}$—with full rank—and $\tau_{\theta,0}^{\text{full}}$ degrees of freedom. The degrees-of-freedom parameter $\tau_{\theta,0}^{\text{full}}$ is chosen to be equal to the number of periods in the training sample, which is 44 after accounting for initialization. The scale matrix $\bar{Q}_\theta^{\text{full}}$ is chosen to be proportional to the covariance matrix of the time-invariant estimate of the vector of VAR parameters in the training sample. The constant-of-proportionality is calibrated to 4.4E-4 multiplied with the degrees of freedom, where the number 4.4E-4 is calibrated to imply the same amount of time variation per parameter as the prior in De Wind and Gambetti (2014).

For the reduced-rank model version, the prior is likewise assumed to be an inverse-Wishart distribution, yet now with singular scale matrix $\bar{Q}_\theta^{\text{red}}$, where $\text{rank}(\bar{Q}_\theta^{\text{red}}) = q_\theta$. The main results of this paper are based on the setting $q_\theta = 7$ and, in addition, the results are shown to be robust along this dimension. The singular scale matrix $\bar{Q}_\theta^{\text{red}}$ is constructed from the non-singular scale matrix $\bar{Q}_\theta^{\text{full}}$ by selecting the q_θ most important eigenvalues, i.e. by replacing the $k - q_\theta$ smallest eigenvalues by zeros in the eigenvalue decomposition. The prior for the reduced-rank model version is supposed to imply the same amount of time variation in the parameters as the prior for the full-rank model version. This requires an upscaling since the sum of the eigenvalues—which measures the amount of time variation—has decreased by the rank reduction. This is exactly offset by multiplying the scale matrix with the sum of all the eigenvalues divided by the sum of the q_θ largest eigenvalues. Moreover, the degrees of freedom are chosen to be equal to $\tau_{\theta,0}^{\text{red}} = \tau_{\theta,0}^{\text{full}} - k + q_\theta$ in order to match exactly the expected sum of the eigenvalues in the reduced-rank model version with the full-rank model version, see De Wind and Gambetti (2014) for details.

3.3.2   Prior settings TV-VAR-SV

The prior settings regarding the time-varying parameters part of the model are exactly the same as in the TV-VAR-NO specification and the prior settings regarding the stochastic volatility part of the model are basically identical to the baseline prior settings of Primiceri (2005) except for one minor difference. As explained in footnote 10, Primiceri (2005) imposes a block-diagonal structure on the covariance matrix of the shocks driving the off-diagonal elements and therefore has to specify separate priors for the various blocks, whereas I do not impose a block-diagonal structure on this matrix and therefore can simply specify one prior for the whole matrix. The strategy for constructing the prior is nevertheless identical, using one plus the dimension as degrees of freedom and using the same constant-of-proportionality.

4   Results

I have estimated the time-varying VAR model (without stochastic volatility, i.e. the TV-VAR-NO spec-
ification) and I have used the proxy identification procedure to identify the structural tax shock.[14] In
this section, I present the results, first on the time variation in the cyclicality of fiscal policy and then on
the time variation in the dynamic effects of unanticipated changes in tax policy. I conclude this section
with various robustness checks, among which the TV-VAR-SV robustness check.

4.1   Time variation in the cyclicality of fiscal policy

In this subsection, I present systematic evidence on the time variation in the cyclicality of fiscal policy.[15]
The time variation in the comovement between the fiscal variables and GDP can be determined by the
time-varying covariance matrix implied by the time variation in the VAR coefficients.[16] However, I am
interested in the *response* of the fiscal variables to the fluctuations in the economy, except those fluctu-
ations that are originally caused by shocks in fiscal policy—after all, the causal relationship would then
have been the other way around. The key object for measuring the cyclicality of fiscal policy is therefore
the conditional covariance matrix between the fiscal variables and GDP conditional on only non-policy
shocks. I have used the proxy identification procedure to identify the structural tax shock and with one
additional identification assumption (that government expenditure cannot respond contemporaneously
to non-policy shocks) I can also identify the structural government expenditure shock. The rest of the
fluctuations in the economy are assumed to be caused by non-policy shocks, on which the rest of this
subsection is conditioned.

Regarding US tax policy, figure 6 plots the time-varying covariance between tax revenues and GDP
conditional on non-policy shocks. The figure shows that the conditional covariance between tax revenues
and GDP has been positive during the post World War II period. The positive conditional covariance

---

[14]The Bayesian estimation is based on four Markov chains of $20,000$ draws each and the first 50% of each chain is thrown
away as burn-in period. For the current application, this is sufficient for the Markov chains to converge to the posterior
distribution, which I have confirmed by comparing the results of various Markov chains that have started from different
initial conditions. The results are constructed from the median time paths of the VAR parameters.

[15]The VAR model does not contain sufficient information to distinguish between automatic stabilizers and discretionary
policy, and the results in this subsection are therefore for the two policy components together.

[16]The time-varying covariance matrix can be obtained from a simulation or by a so-called local approximation (that
approximates the covariance matrix at each point in time by imposing that the VAR coefficients will remain constant in
the future). Here, the local approximation is used as in most time-varying VAR papers.

implies that tax policy including both automatic stabilizers as well as discretionary tax policy has been countercyclical (recall that the VAR model does not contain sufficient information to distinguish between the two policy components). The figure also shows that the countercyclicality has become stronger over time with a peak around the turn of the millennium. Note that the 2007–2008 financial crisis and the subsequent stimulus packages are not included in the sample period, so that we might expect another peak.[17] I have also calculated the time-varying correlation between tax revenues and GDP as well as the coefficient of the regression of tax revenues on GDP, both conditional on non-policy shocks. These statistics are closely related to the conditional covariance between tax revenues and GDP and support the exact same story, as shown in figure 7.

Regarding US spending policy, figure 8 plots the time-varying covariance between government expenditures and GDP conditional on non-policy shocks. The figure shows that spending policy used to be procyclical and has become countercyclical after the beginning of the 1990s. This result is consistent with the sample-split results of Auerbach (2008), who shows that fiscal policy has become (more) countercyclical after the introduction of the Budget Enforcement Act of 1990, both regarding spending as well as tax policy. I have also calculated the time-varying conditional correlation and regression coefficient between government expenditures and GDP, and these statistics support the exact same story, as is clear from figure 9.

4.2   Time variation in the dynamic effects of unanticipated changes in tax policy

Given the increased use of countercyclical fiscal policy, it is important to understand the dynamic effects of changes in fiscal policy. In this subsection, I present results on the dynamic effects of unanticipated changes in tax policy, and in particular how the effects have changed over time. The analysis is based on the time-varying impulse response function of the structural tax shock, where the structural tax shock is identified as in Mertens and Ravn (2013b). The distinctive difference with Mertens and Ravn (2013b) is the time variation in the impulse response function that arises from the time variation in the reduced-form parameters of the VAR model, i.e. I extend their analysis by allowing for time variation in the

---

[17]The time-varying VAR model setup is designed to capture structural changes and is not well-suited to capture changes at business-cycle frequencies, which is confirmed by figure 6 that displays no interesting time variation in the shaded NBER recessions. Nevertheless, even though the 2007–2008 financial crisis and the subsequent stimulus packages are business-cycle phenomena, we might expect another peak given that the stimulus packages have been rather large in absolute terms.

reduced-form parameters.[18]

Figure 10 plots the impulse response function of the structural tax shock for various points in time. The response on impact is, by construction, time-invariant and, as a matter of fact, somewhat larger than in Mertens and Ravn (2013b), at least for GDP. From the second period onwards, the time variation in the reduced-form parameters of the VAR model enters the impulse response function of the structural tax shock.

In this paper, I focus on the response of GDP and, for convenience, I have magnified the response of GDP in figure 11. The figure produces two interesting results. First, there has been a permanent decline in the tax multiplier, from more than four at the beginning of the sample period to around three at the end of the sample period. Despite the permanent decline, the estimated tax multiplier is still at the higher end of the range of existing empirical estimates, which is consistent with Mertens and Ravn (2013b). Second, the response of the economy has become much faster, reaching its peak response after six quarters at the beginning of the sample period as opposed to four quarters at the end of the sample period. Correspondingly, the effects of the structural tax shock also fade out somewhat faster.

The permanent decline in the tax multiplier and faster response of the economy found in this paper are consistent with the sample-split results of Mertens and Ravn (2013b). I have replicated their impulse response function for the pre-1980 period and the post-1980 period, see section 2.4 and in particular figure 3. The economy reaches its peak response after six quarters in the early sub-sample as opposed to four quarters in the later sub-sample and also the tax multiplier has declined considerably, just like with my time-varying VAR.

Finally, it should be noted that the structural tax shock is identified based on a narrative series of unanticipated changes in tax policy that are motivated either as ideological or as arising from inherited debt concerns, see Romer and Romer (2009) and Mertens and Ravn (2013b) for more details. Therefore, the results presented in this subsection apply to tax policy that is motivated independent of the state of the business cycle. In contrast, the results in the previous subsection apply to tax policy that is motivated as response to fluctuations in the economy. It is important to be cautious with this distinction. Of course, it is quite possible that the permanent decline in the tax multiplier and faster response of the economy found in this subsection carry over to tax policy that is motivated by the state of the business cycle (e.g. stimulus in downturn), but this is still undetermined and should be investigated, which is left for future

---

[18]The identification step itself is time-invariant as in Mertens and Ravn (2013b). There are too few nonzero elements in the narrative series of unanticipated tax changes to reliably introduce time variation in the identification step.

research.

4.3   Robustness checks

I have checked for robustness along various dimensions and the main results of this paper are very robust. In particular, I have generated results with the TV-VAR-SV specification (i.e. the model specification with stochastic volatility), investigated the influence of the settings regarding the covariance matrix of the shocks driving the time-varying parameters (in particular, the chosen rank and the prior settings), and estimated a model version with only two lags. For space considerations, I only present figures on the time variation in the response of GDP to the structural tax shock and the time-varying covariance between tax revenues and GDP conditional on non-policy shocks. The various robustness checks are presented in turn.[19]

4.3.1   Stochastic volatility

Allowing for time variation in the covariance matrix of the reduced-form shocks (i.e. stochastic volatility) does not contribute directly to the time variation in the impulse response function of the structural tax shock. Nevertheless, the TV-VAR-SV specification serves as an important robustness check as the incorporation of stochastic volatility could possibly matter in an indirect manner. In particular, many model properties depend both on the reduced-form parameters of the VAR model and the covariance matrix of the reduced-form shocks, so that wrongly leaving out time variation in the latter may lead to spurious time variation in the former (or the other way around).[20] The figures presented in this subsection suggest that this is *not* a problem for the current application, i.e. the TV-VAR-NO and TV-VAR-SV results are consistent with each other.

Figure 12 plots the response of GDP to the structural tax shock for various points in time. The

---

[19]For all robustness checks, the Bayesian estimation is based on four Markov chains of $20,000$ draws each and the first 50% of each chain is thrown away as burn-in period. This is sufficient for the Markov chains to converge to the posterior distribution, which I have confirmed by comparing the results of various Markov chains that have started from different initial conditions. The results are constructed from the median time paths of the VAR parameters.

[20]For example, consider the simplest possible AR model, $y_t = \varrho y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N\left(0, \sigma^2\right)$. The unconditional variance of y_t depends on both model parameters, that is $\text{var}\left(y_t\right) = \frac{\sigma^2}{1-\varrho^2}$. Suppose that the variance of the shocks σ^2 has increased over time. Then, if we estimate a time-varying version of the AR(1) allowing only for time variation in the autoregressive parameter ϱ, we will find that the persistence of y_t has increased over time, which is a spurious result. Allowing for time variation in both model parameters helps to avoid such spurious results. This argument was put forward by Stock (2001).

permanent decline in the tax multiplier and faster response of the economy found in this paper are confirmed by the TV-VAR-SV specification, at least in qualitative terms. Yet, in quantitative terms, there is somewhat less time variation, which is not surprising given that some of the time variation that was previously captured by the time-varying parameters will now be absorbed by the stochastic volatility (which will not be reflected in the impulse response function of the structural tax shock), see De Wind (2013) for an extensive discussion.

The left panel of figure 13 plots the time-varying covariance between tax revenues and GDP conditional on non-policy shocks. The conditional covariance depends directly on the stochastic volatility and is, in fact, dominated by movements in the covariance matrix of the reduced-form shocks (high-frequency movements as well as the Great Moderation). To filter out the high-frequency movements, I have constructed a two-sided moving average with ten quarters on each side, as plotted in the right panel of figure 13. It seems that the results are not inconsistent with the TV-VAR-NO specification.

4.3.2   Number of factors

I have no formal procedure for choosing how many factors to include in the reduced-rank model. Therefore, it is important to check for robustness along this dimension. I have generated results for all possible ranks up to ten as well as for the full-rank model. For a selection of the ranks, I have plotted the response of GDP to the structural tax shock for various points in time (figure 14) and the time-varying covariance between tax revenues and GDP conditional on non-policy shocks (figure 15).

There are 39 parameters in the trivariate VAR with four lags and constant, yet it is evident from the figures that much less underlying factors are needed to capture the bulk of time variation that is present in the full-rank model. As few as three factors are sufficient to replicate the time patterns found with the full-rank model, albeit only qualitatively. With five factors, the differences are already quite small in quantitative terms. And from seven factors onwards, the differences are hardly distinguishable. It seems that including seven factors is the sweet spot for the current application, and was therefore chosen as the benchmark case.

4.3.3   Prior tightness and informativeness

One of the conclusions of De Wind (2013) was that the inverse-Wishart prior on the covariance matrix of the shocks driving the time-varying parameters has much influence on the estimated amount of time variation. Checking for robustness along this dimension is therefore important. I have played around

with the constant-of-proportionality that was used to scale the scale matrix and also with the degrees-of-freedom parameter.

Regarding the constant-of-proportionality, I have generated results with constants-of-proportionality that are ten times tighter ($\times 0.1$), five times tighter ($\times 0.2$), two times tighter ($\times 0.5$), two times looser ($\times 2$), and five times looser ($\times 5$). As before, I have plotted the response of GDP to the structural tax shock for various points in time and the time-varying covariance between tax revenues and GDP conditional on non-policy shocks in figures 16 and 17, respectively. The constant-of-proportionality has some influence on the estimated amount of time variation, yet the main results even survive the ten times tighter prior, at least in qualitative terms.

Regarding the degrees-of-freedom parameter, I have played around with a less informative prior (less degrees of freedom) as well as with a more informative prior (more degrees of freedom). Figure 18 plots the results, the response of GDP to the structural tax shock for various points in time in the left column and the time-varying covariance between tax revenues and GDP conditional on non-policy shocks in the right column. The degrees-of-freedom parameter has some influence on the estimated amount of time variation, yet the main results are robust, at least in qualitative terms.

4.3.4 Two lags

As a final robustness check, I have estimated a model version with only two lags.[21] A confirmation from this model version is an important piece of evidence that my results are robust, given that the risk of overfitting is much lower with only 21 parameters in the trivariate VAR. Besides, it is interesting to see whether four lags are really needed in the standard fiscal VAR analyzed here, which was advocated by Blanchard and Perotti (2002).

As before, I have plotted the response of GDP to the structural tax shock for various points in time and the time-varying covariance between tax revenues and GDP conditional on non-policy shocks in figures 19 and 20, respectively. It is questionable whether four lags are really needed in the standard fiscal VAR analyzed here. In particular, the shape of the dynamic response of GDP to the structural tax shock is very similar as in the model version with four lags. The impact response of GDP is estimated to be slightly lower, yet the estimated amount of time variation is very close, quantitatively.

---

[21]I have re-calibrated the prior on the covariance matrix of the shocks driving the time-varying parameters to imply the same amount of time variation per parameter as in the model version with four lags.

5   Concluding remarks

Using a structural vector autoregression with time-varying parameters, I have analyzed to what extent the dynamic effects of unanticipated changes in tax policy have changed structurally over the post World War II period in the United States, and I have also produced results on the time variation in the cyclicality of fiscal policy.

Regarding the time variation in the dynamic effects of unanticipated changes in tax policy, the analysis is based on the time-varying impulse response function of the structural tax shock, where the structural tax shock is identified as in Mertens and Ravn (2013b) using the same (time-invariant) identification assumptions and narrative series of unanticipated tax changes. All the time variation in the impulse response function arises from the time variation in the reduced-form parameters of the VAR model, while the identification step itself is time-invariant. Clearly, this is unfortunate since this causes the response on impact to be time-invariant.[22] It is left for future research to introduce time variation in the identification step, which will be challenging given that there are only a few nonzero elements in the narrative series of unanticipated tax changes. Potential routes are combining multiple proxies, combining the proxy identification procedure with other identification procedures, and exploiting the informational content of the zero elements in the proxy (of which there are many).

The estimated time variation points to a permanent decline in the tax multiplier as well as a faster response of the economy. It should be noted, however, that the permanent decline in the tax multiplier does *not* suggest that tax cuts are a bad policy instrument for stimulating the economy, as the estimated tax multiplier is still at the higher end of the range of existing empirical estimates (which is consistent with Mertens and Ravn (2013b), whose identification strategy I have followed). Moreover, the time-varying VAR model setup is designed to capture structural change in the effects of fiscal policy, and is not well-suited to capture the business-cycle dependence of the effects. There is a large literature that argues that fiscal multipliers are larger when the economy is in recession. It would be very useful to model both types of time variation in one unified framework.

Finally, regarding the time variation in the cyclicality of fiscal policy, the estimated time variation

---

[22]At the same time, given that all the time variation in the impulse response function arises from the time variation in the reduced-form parameters of the VAR model, it is likely that the direction of the time variation carries over to other identification assumptions. In fact, suppose that the impact response of GDP would have been half as large, the estimated time variation would have pointed in the exact same direction. Of course, this example cannot replace a proper analysis of other identification assumptions.

suggests that fiscal policy has become more countercyclical over time. In particular, spending policy used to be procyclical and has become countercyclical after the beginning of the 1990s (around the introduction of the Budget Enforcement Act of 1990), whereas tax policy already used to be countercyclical and has become even more countercyclical over time. These results are consistent with the sample-split results of Auerbach (2008).

A Derivation of β_{TT}

In this appendix, I provide a derivation for the β_{TT} expression that was given in the main text for the scalar case. Actually, I provide a derivation for the more general $\beta_{TT}\beta'_{TT}$ expression that was given by Mertens and Ravn (2013a), which in the scalar case can in turn be simplified to the β_{TT} expression.

My derivation starts with the set of restrictions implied by

$$\Sigma = BB' \tag{17}$$

which, with the same partitioning as in the main text, can be written as

$$\begin{pmatrix} \Sigma_{TT} & \Sigma_{TO} \\ \Sigma_{OT} & \Sigma_{OO} \end{pmatrix} = \begin{pmatrix} \beta_{TT} & \beta_{TO} \\ \beta_{OT} & \beta_{OO} \end{pmatrix} \begin{pmatrix} \beta'_{TT} & \beta'_{OT} \\ \beta'_{TO} & \beta'_{OO} \end{pmatrix} \tag{18}$$

which in turn can be expanded to the following set of restrictions

$$\Sigma_{TT} = \beta_{TT}\beta'_{TT} + \beta_{TO}\beta'_{TO} \tag{19a}$$

$$\Sigma_{OT} = \beta_{OT}\beta'_{TT} + \beta_{OO}\beta'_{TO} \tag{19b}$$

$$\Sigma_{OO} = \beta_{OT}\beta'_{OT} + \beta_{OO}\beta'_{OO} \tag{19c}$$

where I have omitted the redundant equation for Σ_{TO}, i.e. the equations for Σ_{TO} and Σ_{OT} are just the transpose of each other.

The next ingredient for the derivation is the term $\beta_{OT}\beta_{TT}^{-1}$, for which we already have an expression in terms of quantities that we can estimate. This term, from now onwards denoted by Z, can be used to substitute out β_{OT} from the above set of restrictions, yielding

$$\Sigma_{TT} = \beta_{TT}\beta'_{TT} + \beta_{TO}\beta'_{TO} \tag{20a}$$

$$\Sigma_{OT} = Z\beta_{TT}\beta'_{TT} + \beta_{OO}\beta'_{TO} \tag{20b}$$

$$\Sigma_{OO} = Z\beta_{TT}\beta'_{TT}Z' + \beta_{OO}\beta'_{OO} \tag{20c}$$

Now equation (20a) is the logical starting point for solving β_{TT}, which can be rewritten as

$$\beta_{TT}\beta'_{TT} = \Sigma_{TT} - \beta_{TO}\beta'_{TO} \tag{21}$$

which suggests that we need to uncover an expression for $\beta_{TO}\beta'_{TO}$. By taking the difference between equation (20b) and Z times equation (20a), we get

$$\Sigma_{OT} - Z\Sigma_{TT} = \beta_{OO}\beta'_{TO} - Z\beta_{TO}\beta'_{TO} = (\beta_{OO} - Z\beta_{TO})\beta'_{TO} \tag{22}$$

from which we can derive the following expression

$$\beta_{TO}\beta'_{TO} = (\Sigma_{OT} - Z\Sigma_{TT})' \left((\beta_{OO} - Z\beta_{TO})(\beta_{OO} - Z\beta_{TO})'\right)^{-1} (\Sigma_{OT} - Z\Sigma_{TT}) \tag{23}$$

where the inverse term in the middle still consists of quantities that are unknown. We can, however, uncover an expression for the inverse term in the middle by subtracting equation (20b) times Z' and Z times the transpose of equation (20b) from Z times equation (20a) times Z' and adding equation (20c), which yields the following expression

$$Z\Sigma_{TT}Z' - \Sigma_{OT}Z' - Z\Sigma'_{OT} + \Sigma_{OO} = Z\beta_{TO}\beta'_{TO}Z' - \beta_{OO}\beta'_{TO}Z' - Z\beta_{TO}\beta'_{OO} + \beta_{OO}\beta'_{OO} \tag{24}$$

which can be rewritten by completing the squares, yielding

$$Z\Sigma_{TT}Z' - \Sigma_{OT}Z' - Z\Sigma'_{OT} + \Sigma_{OO} = (\beta_{OO} - Z\beta_{TO})(\beta_{OO} - Z\beta_{TO})' \tag{25}$$

Putting all the steps together, we now have an expression for the inverse term in the middle of equation (23) which we can use in turn to substitute out $\beta_{TO}\beta'_{TO}$ from equation (21). This completes the derivation of the general $\beta_{TT}\beta'_{TT}$ expression that was given by Mertens and Ravn (2013a). In the scalar case, it is possible to simplify the expression by taking the square root, which yields the required expression

$$\beta_{TT} = \sqrt{\Sigma_{TT} - (\Sigma_{OT} - Z\Sigma_{TT})'\, \Gamma^{-1}\, (\Sigma_{OT} - Z\Sigma_{TT})} \tag{26}$$

where $\Gamma = Z\Sigma_{TT}Z' - \Sigma_{OT}Z' - Z\Sigma'_{OT} + \Sigma_{OO}$. Finally, the equation in the main text can be obtained by substituting back $\beta_{OT}\beta_{TT}^{-1}$ for Z.

B Derivation of the first row of B^{-1}

In this appendix, I show how to derive the first row of the inverse impact matrix B^{-1}. The derivation also makes clear that there are no additional identification assumptions needed.

The derivation starts with the block-wise matrix inversion formula, which applied to the inverse impact matrix gives the following expression

$$\begin{pmatrix} \beta_{TT} & \beta_{TO} \\ \beta_{OT} & \beta_{OO} \end{pmatrix}^{-1} =$$

$$\begin{pmatrix} \left(\beta_{TT} - \beta_{TO}\beta_{OO}^{-1}\beta_{OT}\right)^{-1} & -\left(\beta_{TT} - \beta_{TO}\beta_{OO}^{-1}\beta_{OT}\right)^{-1}\beta_{TO}\beta_{OO}^{-1} \\ -\beta_{OO}^{-1}\beta_{OT}\left(\beta_{TT} - \beta_{TO}\beta_{OO}^{-1}\beta_{OT}\right)^{-1} & \beta_{OO}^{-1} + \beta_{OO}^{-1}\beta_{OT}\left(\beta_{TT} - \beta_{TO}\beta_{OO}^{-1}\beta_{OT}\right)^{-1}\beta_{TO}\beta_{OO}^{-1} \end{pmatrix} \tag{27}$$

It turns out that all the blocks of the impact matrix show up in the first row of the block-wise inverted impact matrix. We already have expressions for the two left blocks of the impact matrix, i.e. β_{TT} and β_{OT}, but not for the two right blocks, i.e. β_{TO} and β_{OO}.

It should be noted, however, that although the two right blocks show up in the first row of the block-wise inverted impact matrix, they only show up in the particular combination $\beta_{TO}\beta_{OO}^{-1}$. The next step in the derivation is therefore to uncover an expression for $\beta_{TO}\beta_{OO}^{-1}$. Now decompose this term as

$$\beta_{TO}\beta_{OO}^{-1} = \left(\beta_{OO}\beta_{TO}'\right)'\left(\beta_{OO}\beta_{OO}'\right)^{-1} \tag{28}$$

after which the first and second term on the right-hand side can be uncovered via equation (20b) and (20c), respectively, yielding the following expression

$$\beta_{TO}\beta_{OO}^{-1} = \left(\Sigma_{OT} - Z\beta_{TT}\beta_{TT}'\right)'\left(\Sigma_{OO} - Z\beta_{TT}\beta_{TT}'Z'\right)^{-1} \tag{29}$$

which completes the proof that we can back out all the terms in the first row of the inverse impact matrix without making additional identification assumptions.

Note that, in contrast, it is not possible to back out all the terms in the second row of blocks of the inverse impact matrix since we would need an expression for β_{OO}. Via equation (20c), we can get an expression for $\beta_{OO}\beta_{OO}'$, yet this does not uniquely pin down β_{OO} itself.

References

AUERBACH, A. J. (2008): "Federal Budget Rules: The US Experience," NBER Working Paper No. 14288, National Bureau of Economic Research.

——— (2009): "Implementing the New Fiscal Policy Activism," *The American Economic Review*, 99(2), pp. 543–549.

AUERBACH, A. J., AND Y. GORODNICHENKO (2012): "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4(2), pp. 1–27.

BAUMEISTER, C., AND G. PEERSMAN (2013): "Time-Varying Effects of Oil Supply Shocks on the US Economy," *American Economic Journal: Macroeconomics*, 5(4), pp. 1–28.

BLANCHARD, O., AND R. PEROTTI (2002): "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, 117(4), pp. 1329–1368.

CHRISTIANO, L., M. EICHENBAUM, AND S. REBELO (2011): "When Is the Government Spending Multiplier Large?," *Journal of Political Economy*, 119(1), pp. 78–121.

DE WIND, J. (2013): "Time-varying vector autoregressions and prior influence," unpublished manuscript.

DE WIND, J., AND L. GAMBETTI (2014): "Reduced-rank time-varying vector autoregressions," CPB Discussion Paper No. 270, CPB Netherlands Bureau for Economic Policy Analysis.

KOOP, G., AND D. KOROBILIS (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends(R) in Econometrics*, 3(4), pp. 267–358.

MERTENS, K., AND M. O. RAVN (2013a): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *The American Economic Review*, 103(4), pp. 1212–1247.

——— (2013b): "A reconciliation of SVAR and narrative estimates of tax multipliers," *Journal of Monetary Economics*, forthcoming.

MOUNTFORD, A., AND H. UHLIG (2009): "What Are the Effects of Fiscal Policy Shocks?," *Journal of Applied Econometrics*, 24(6), pp. 960–992.

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72(3), pp. 821–852.

ROMER, C. D., AND D. H. ROMER (2009): "A narrative analysis of postwar tax changes," unpublished manuscript.

SIMS, C. A., AND T. ZHA (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), pp. 949–968.

STOCK, J. H. (2001): "Evolving Post-World War II U.S. Inflation Dynamics: Comment," *NBER Macroeconomics Annual*, 16, pp. 379–387.

STOCK, J. H., AND M. W. WATSON (2008): "What's New in Econometrics: Time Series," NBER Summer Institute, Lecture 7.

———— (2012): "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity*, pp. 81–135.
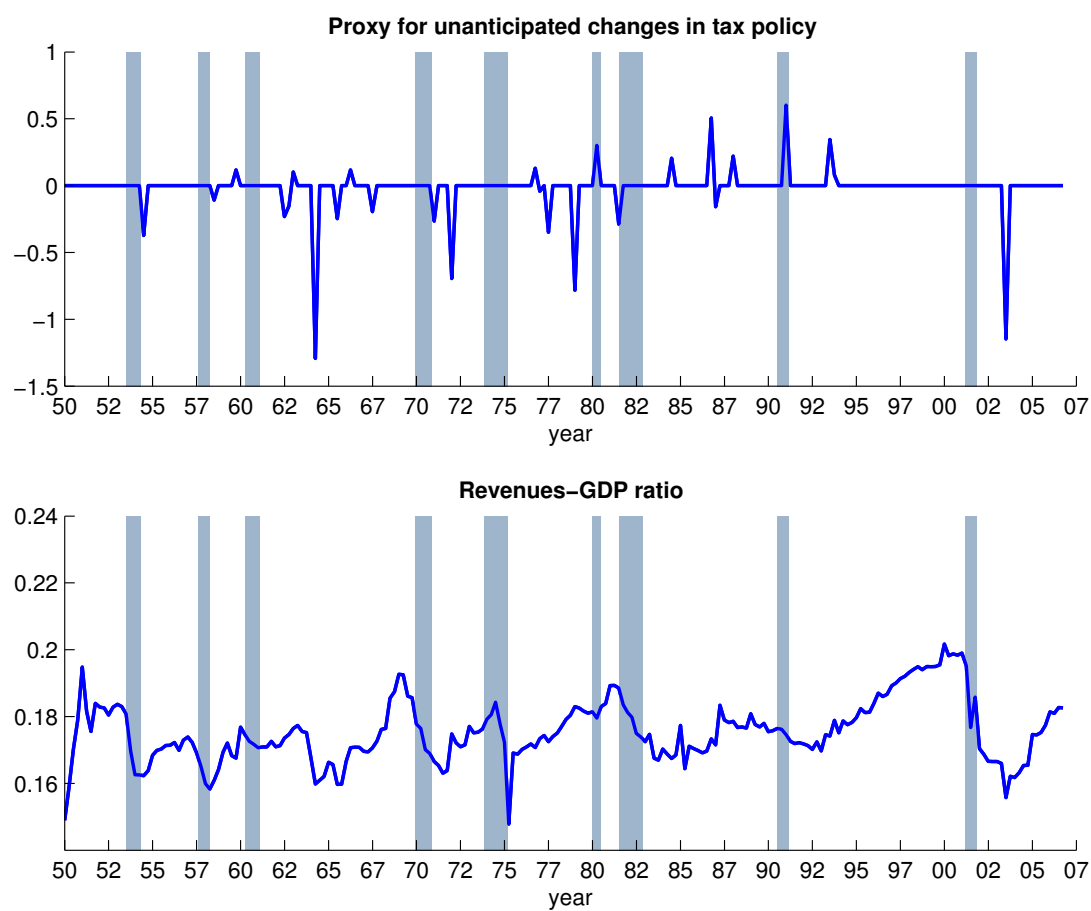
Figure 1: Mertens and Ravn (2013b) proxy for unanticipated changes in tax policy and the revenues-GDP ratio; the NBER recession periods are indicated by the shaded areas
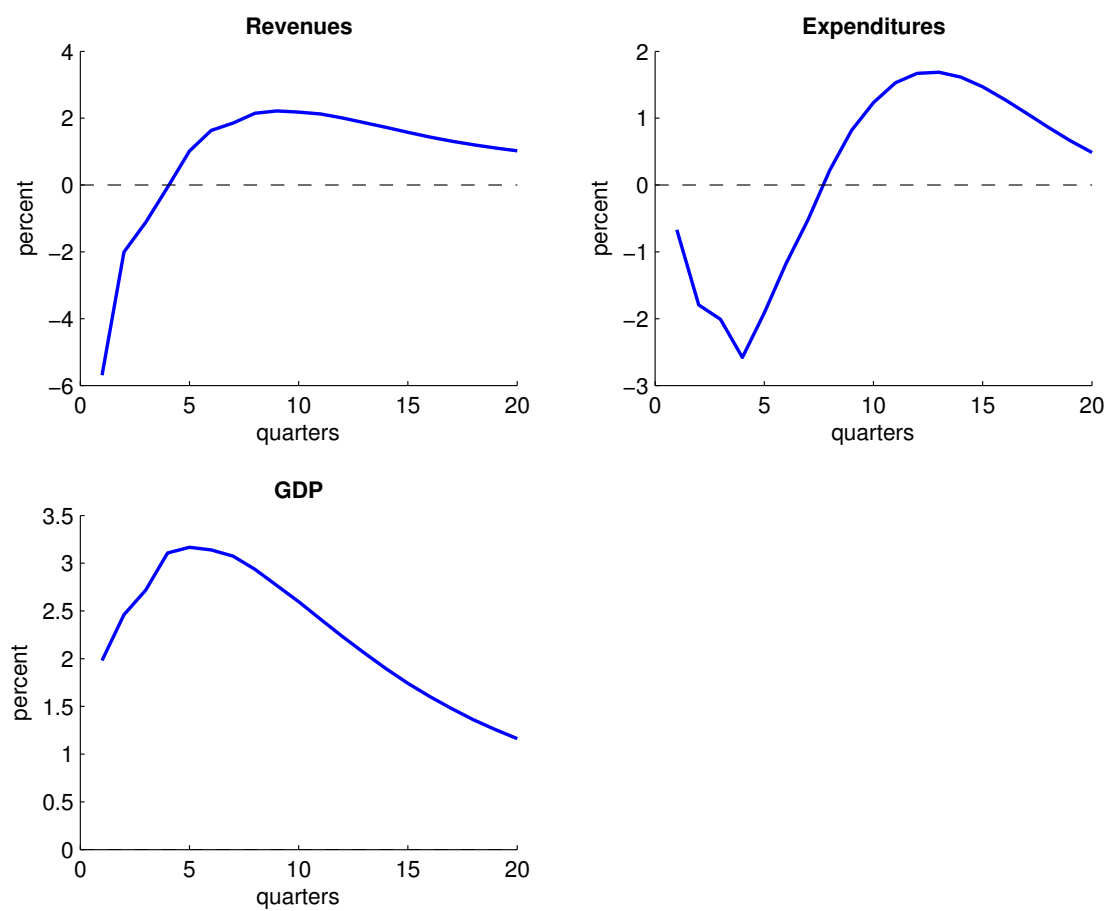
Figure 2: Impulse response function of the structural tax shock for the full sample, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
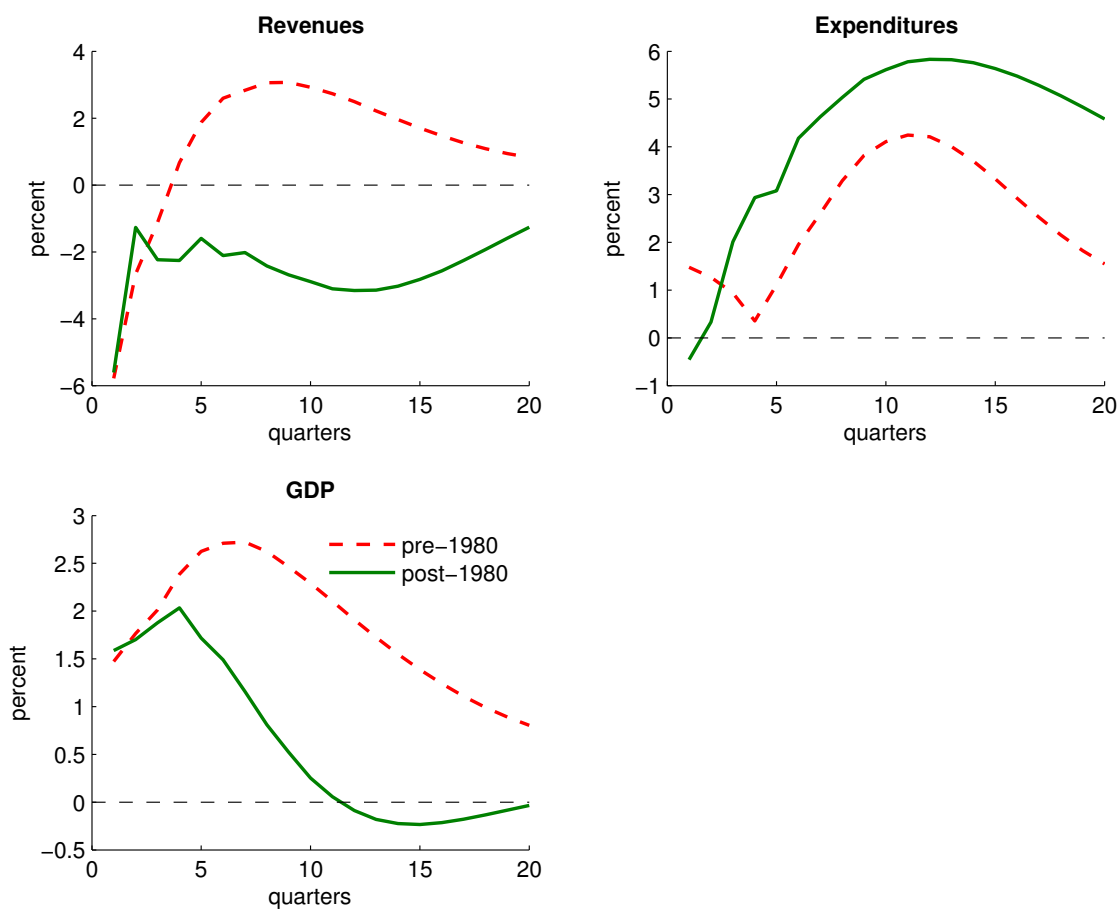
Figure 3: Impulse response function of the structural tax shock for the pre-1980 period and the post-1980 period, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
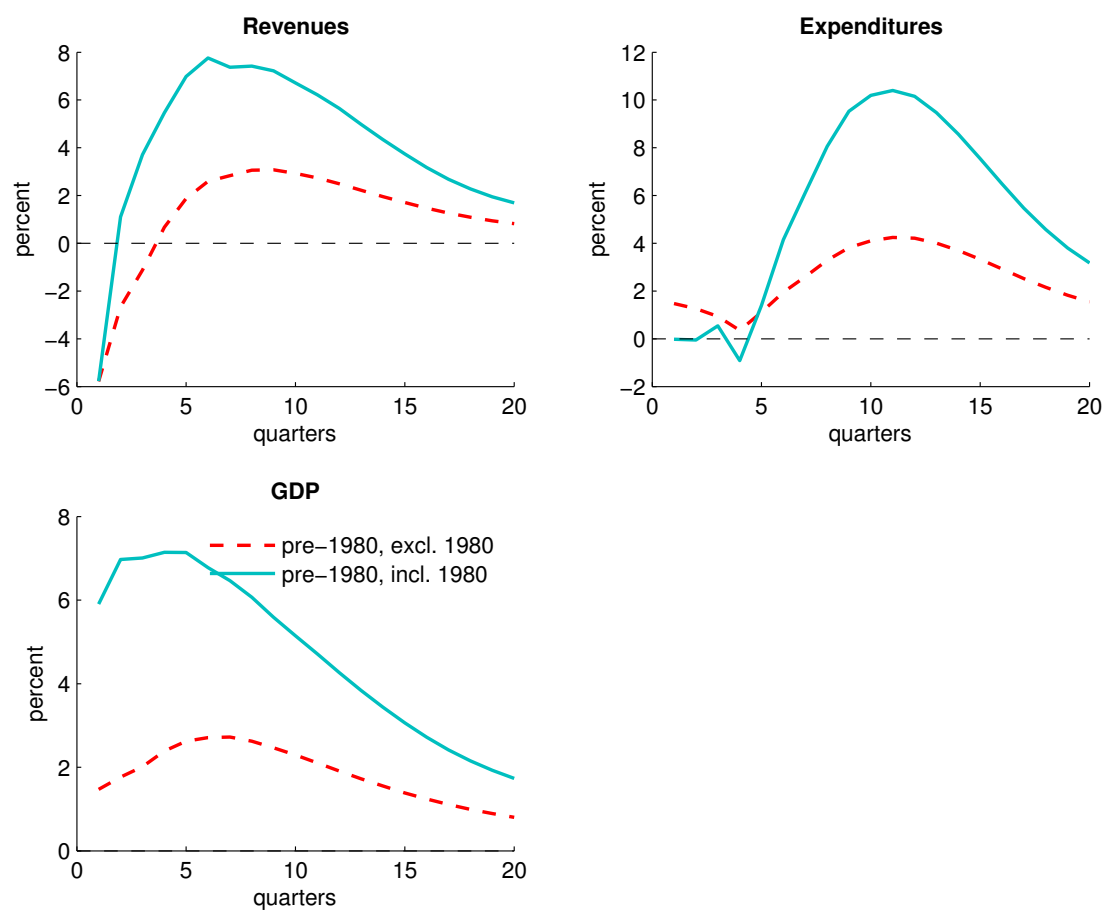
Figure 4: Sample sensitivity of the impulse response function of the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
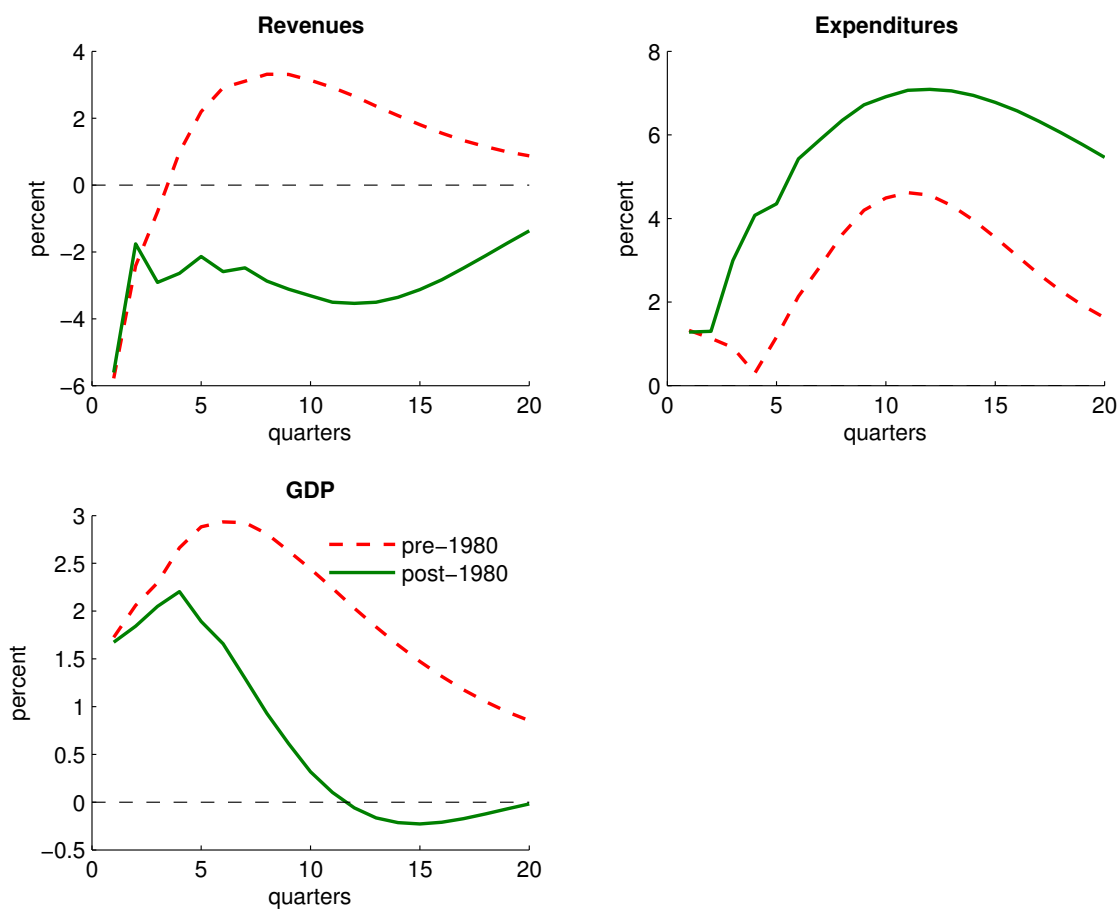
Figure 5: **Impulse response function of the structural tax shock for the pre-1980 period and the post-1980 period, based on the alternative sample-split strategy, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP**
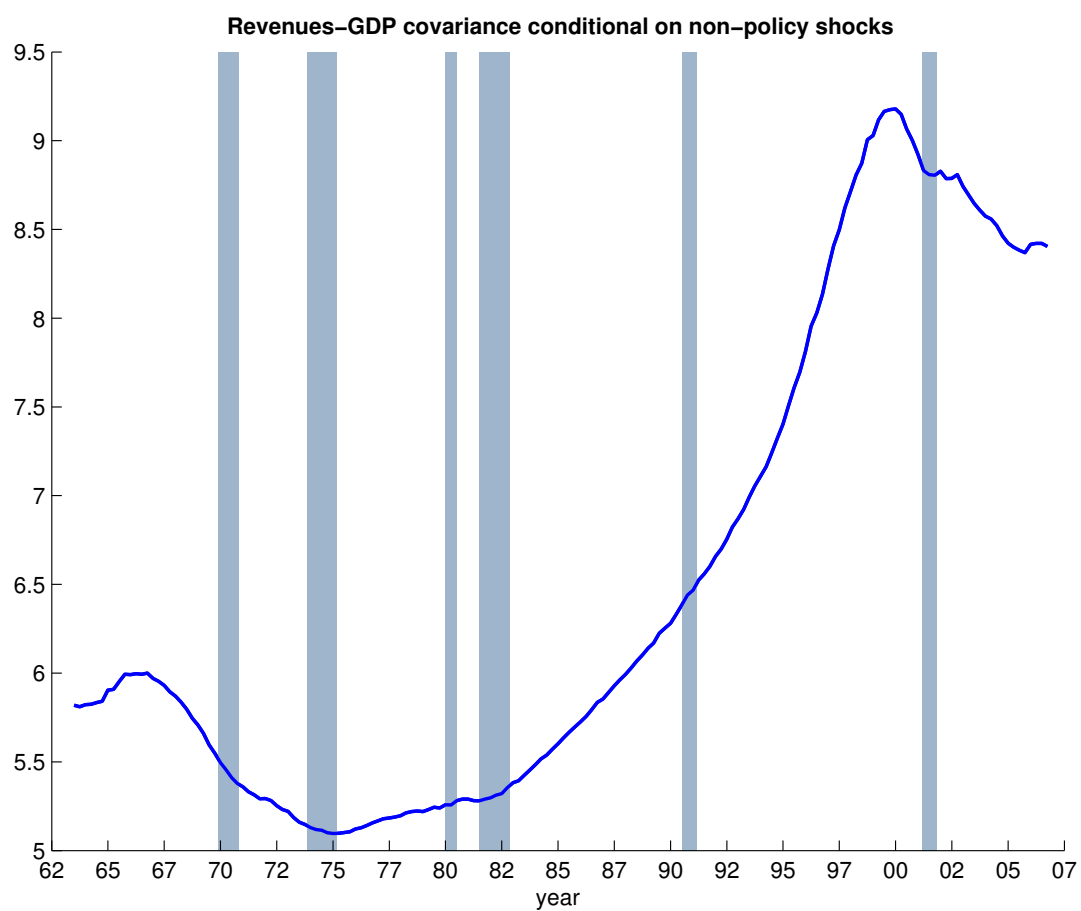
Figure 6: Time-varying covariance between tax revenues and GDP conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas
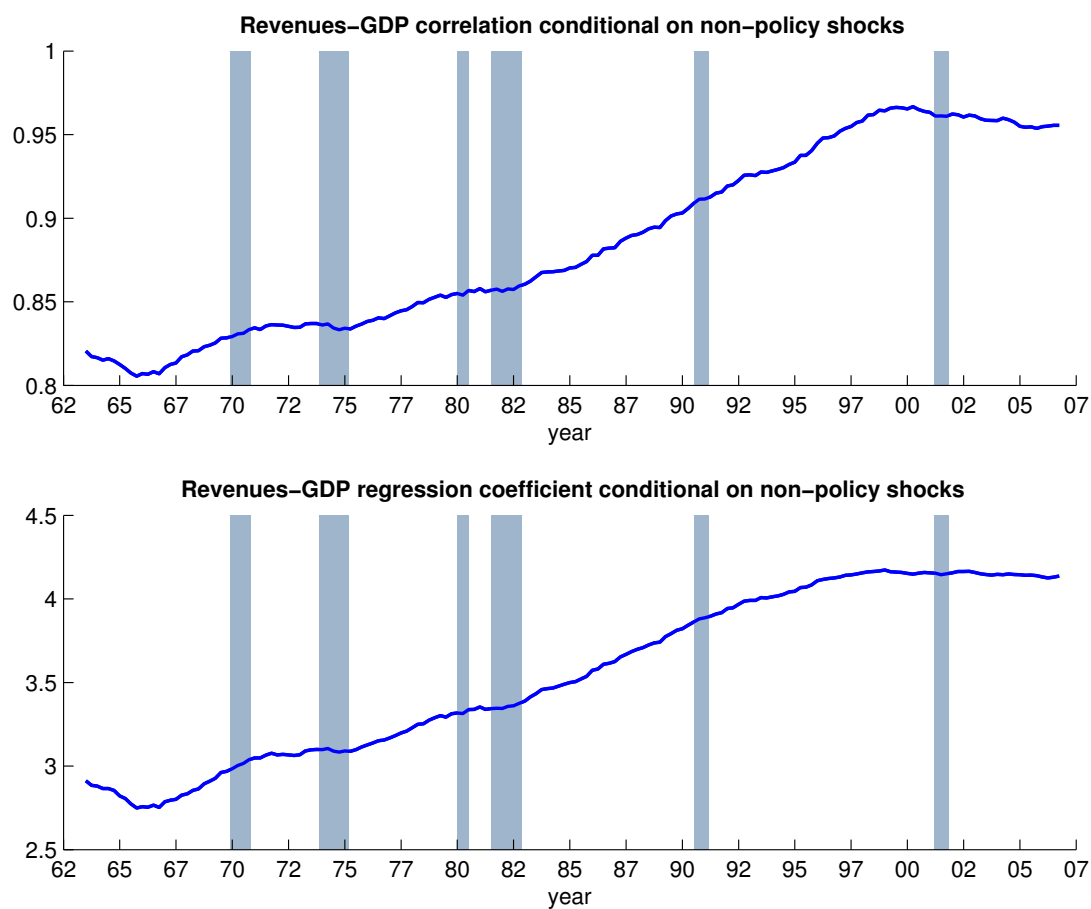
Figure 7: Time-varying correlation between tax revenues and GDP as well as the coefficient of the regression of tax revenues on GDP, both conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas

Figure 8: Time-varying covariance between government expenditures and GDP conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas
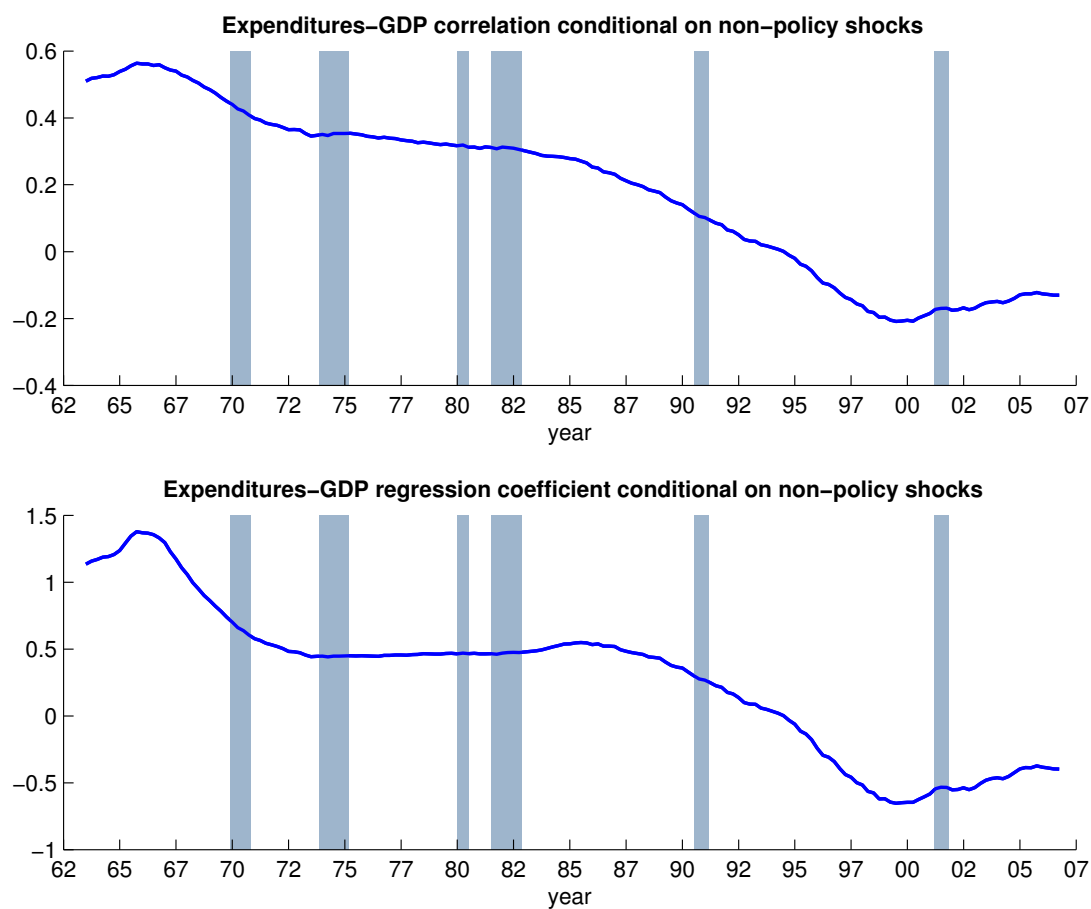
Figure 9: Time-varying correlation between government expenditures and GDP as well as the coefficient of the regression of government expenditures on GDP, both conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas
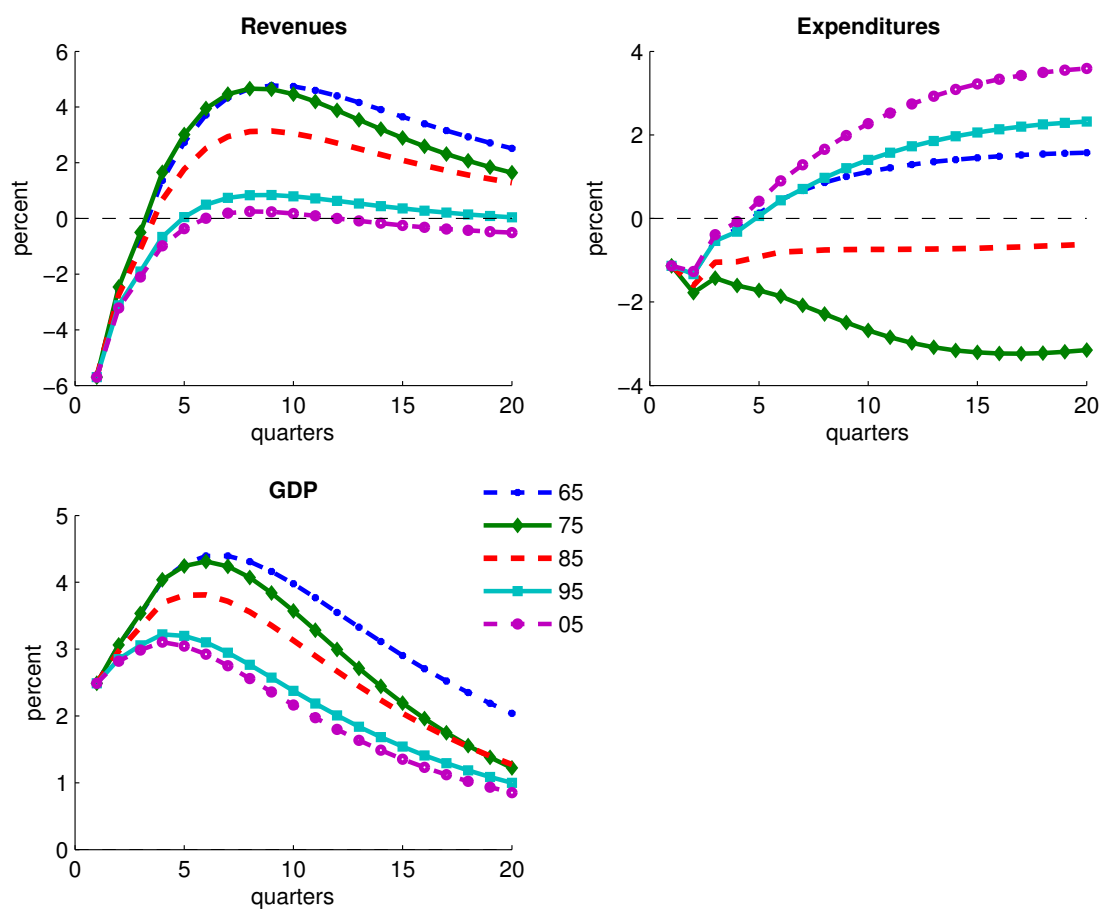
Figure 10: Time variation in the impulse response function of the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
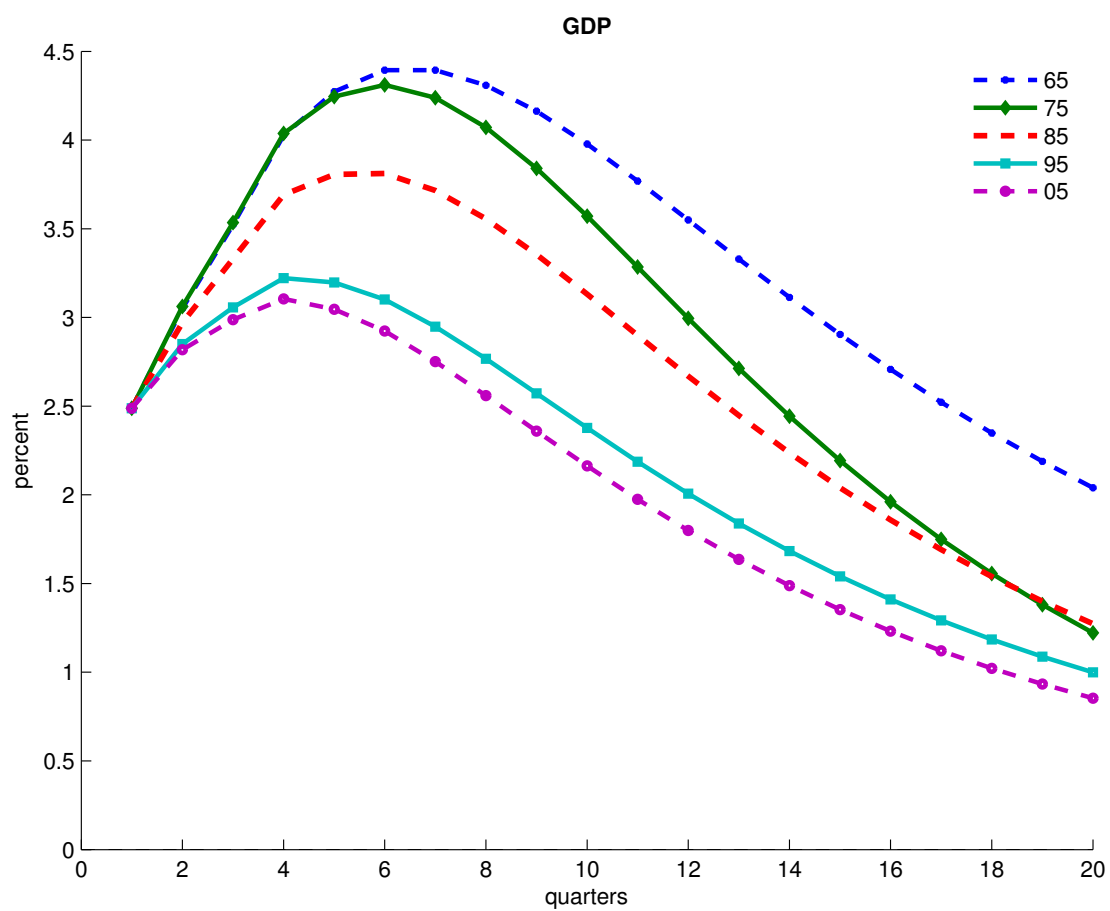
Figure 11: Time variation in the response of GDP to the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
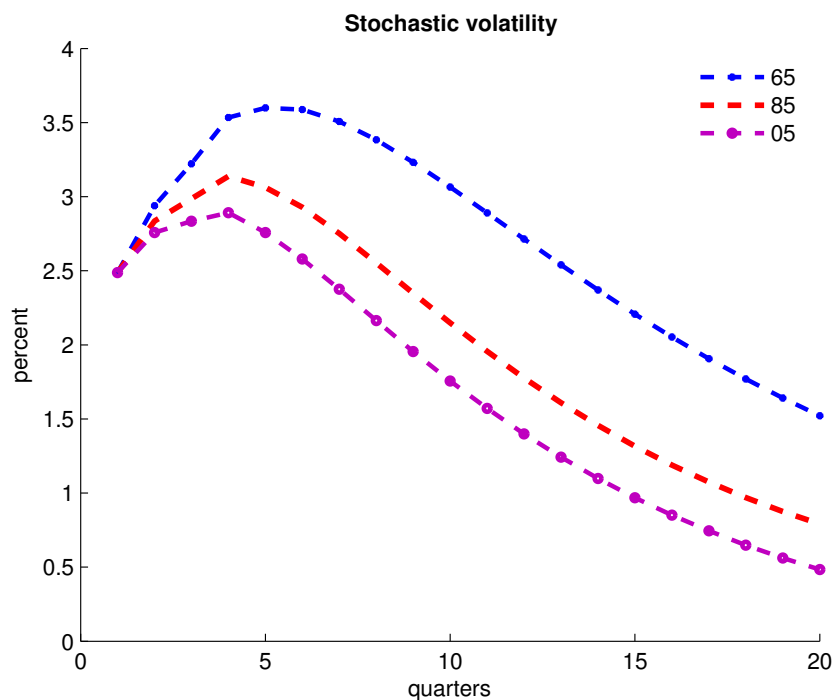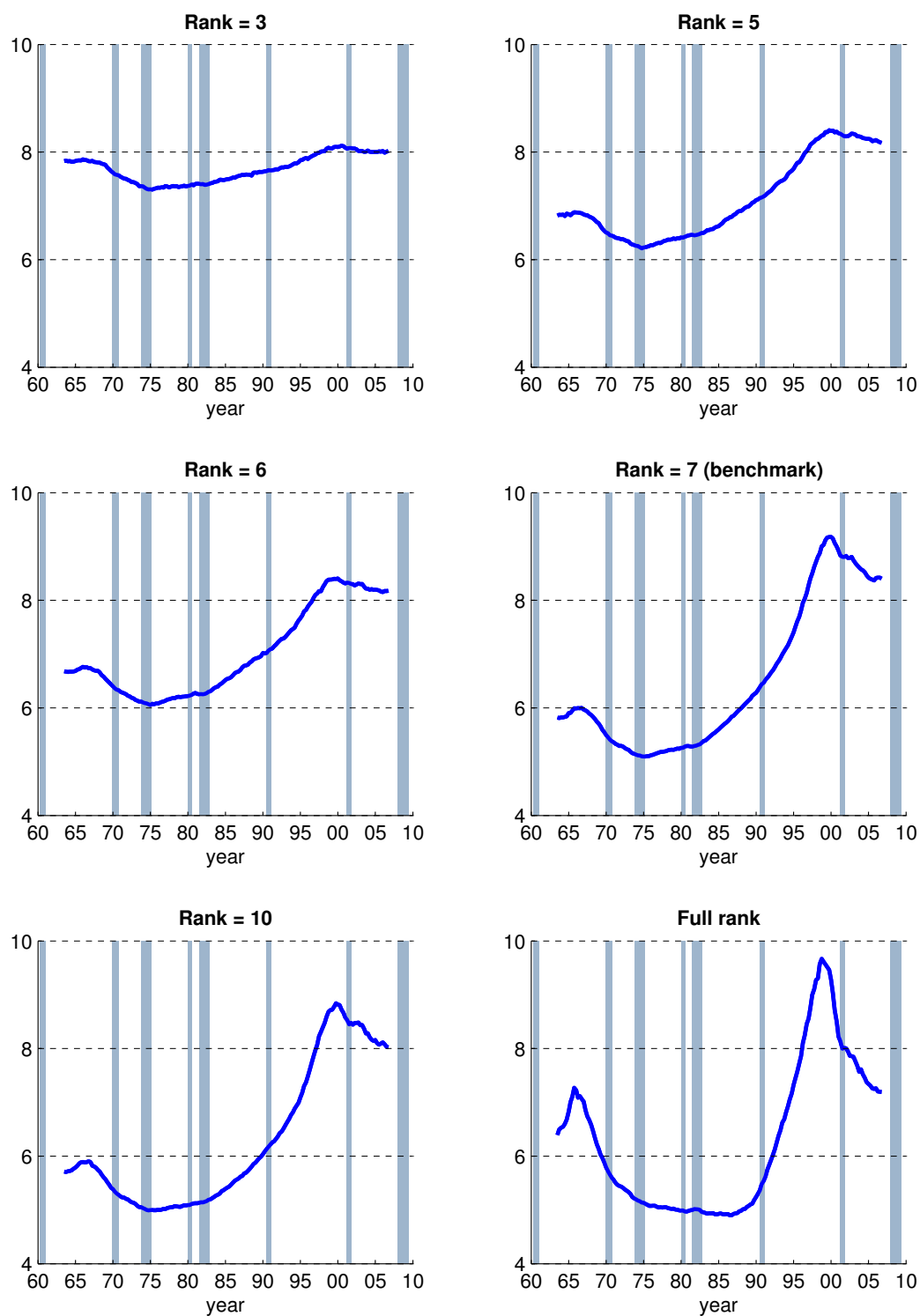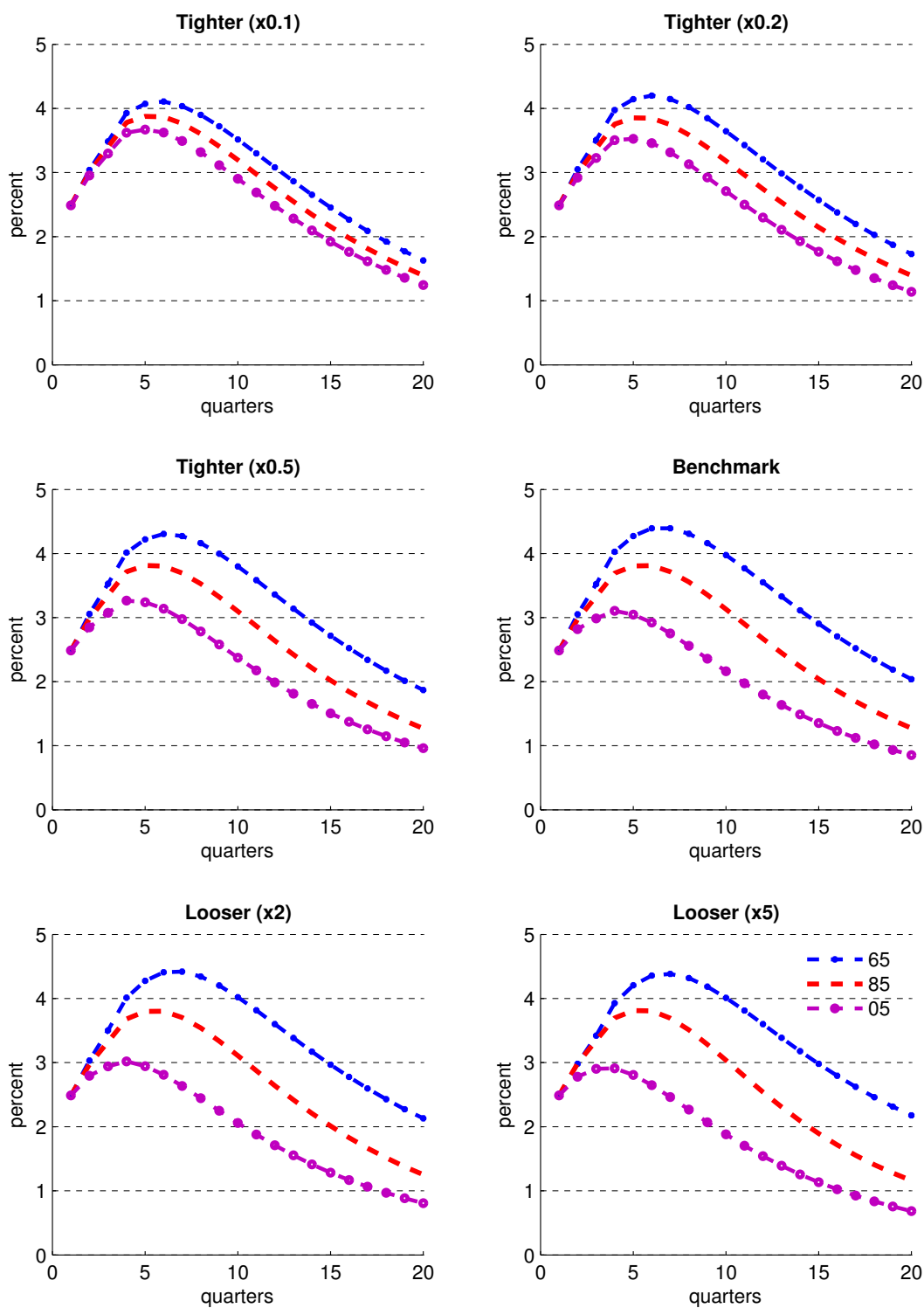
Figure 12: Robustness check, stochastic volatility; time variation in the response of GDP to the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
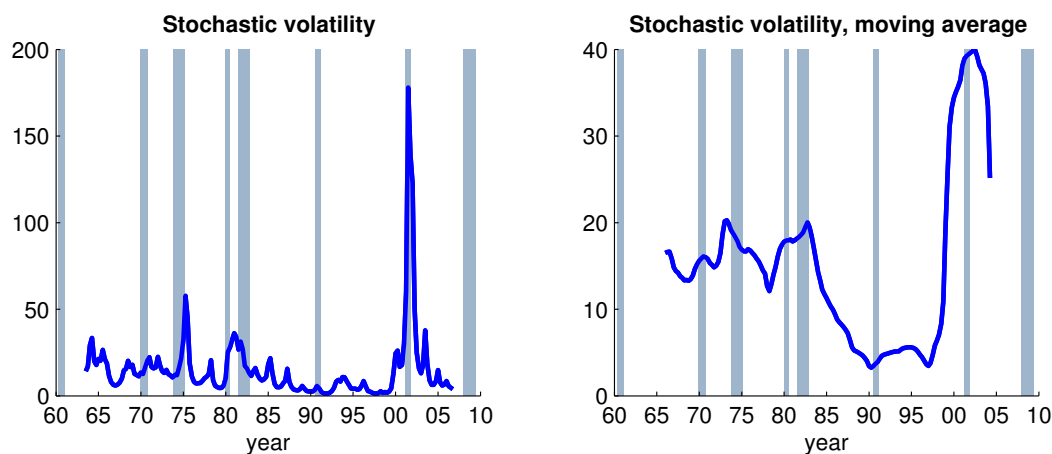
Figure 13: Robustness check, stochastic volatility; time-varying covariance between tax revenues and GDP conditional on non-policy shocks; symmetric two-sided moving average with ten quarters on each side in right panel; the NBER recession periods are indicated by the shaded areas
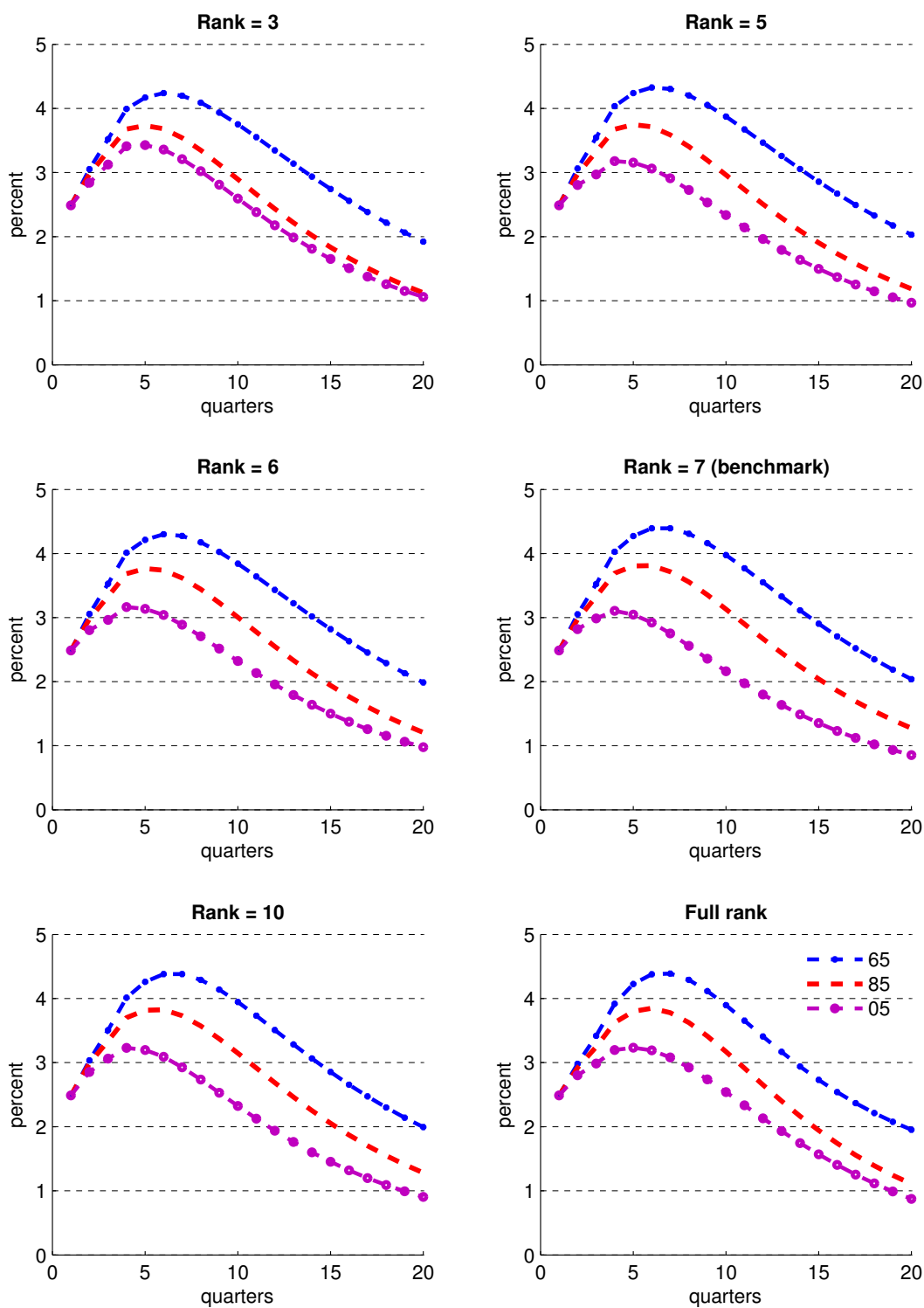
Figure 14: Robustness check, number of factors; time variation in the response of GDP to the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
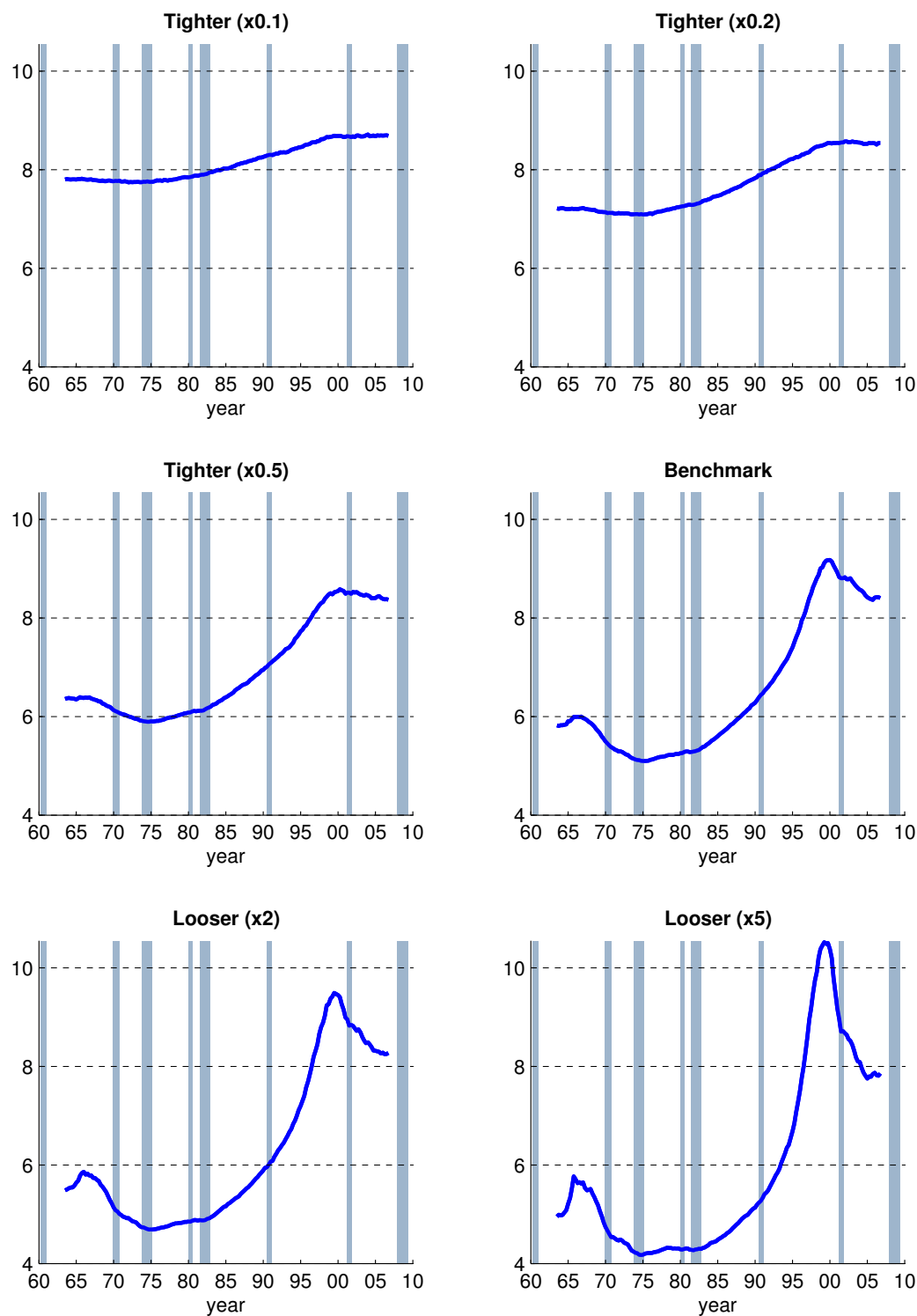
Figure 15: Robustness check, number of factors; time-varying covariance between tax revenues and GDP conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas
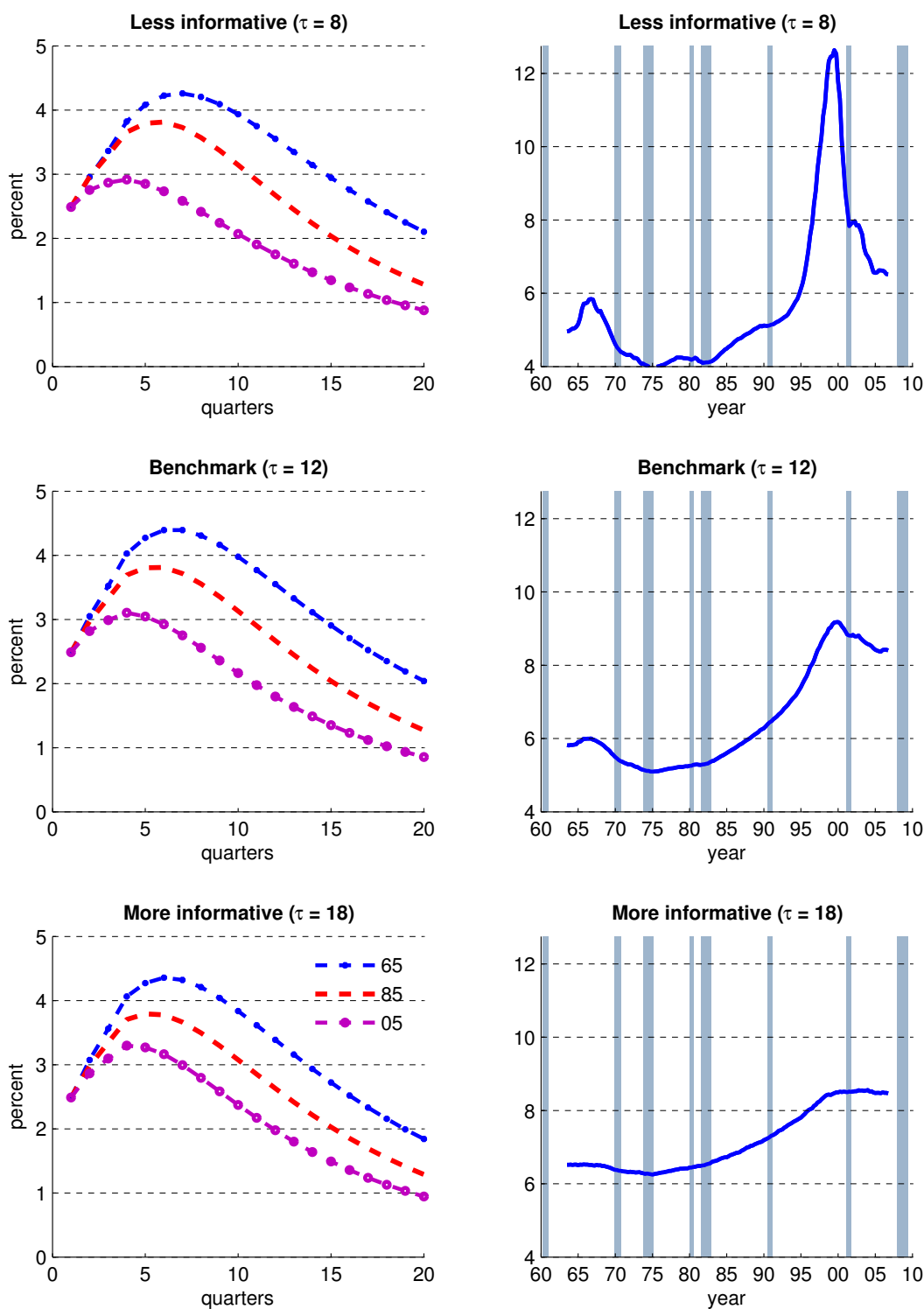
Figure 16: Robustness check, prior tightness; time variation in the response of GDP to the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP

Figure 17: **Robustness check, prior tightness; time-varying covariance between tax revenues and GDP conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas**

Figure 18: Robustness check, prior degrees of freedom; left column: time variation in the response of GDP to the structural tax shock; right column: time-varying covariance between tax revenues and GDP conditional on non-policy shocks
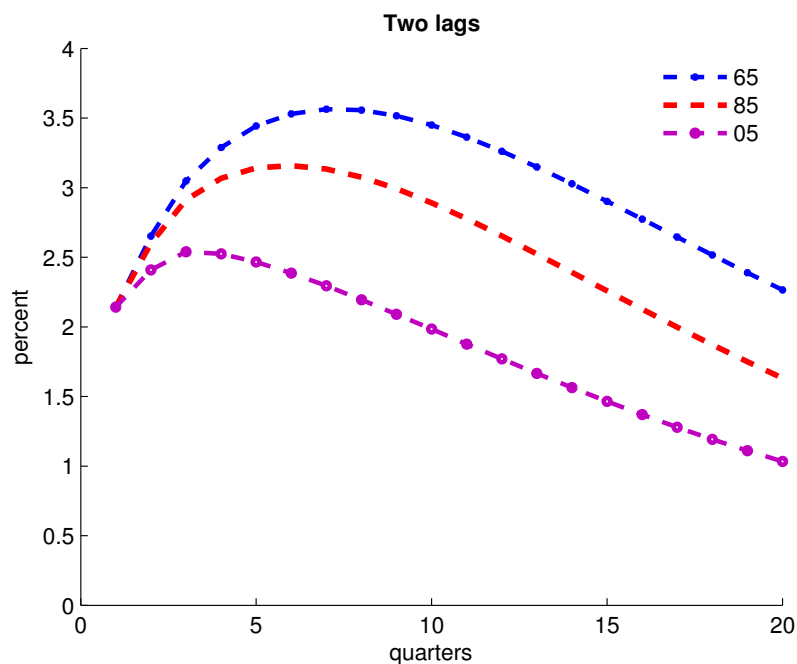
Figure 19: Robustness check, two lags; time variation in the response of GDP to the structural tax shock, impulse scaled such that it amounts to a decrease in tax revenues of one percent of GDP
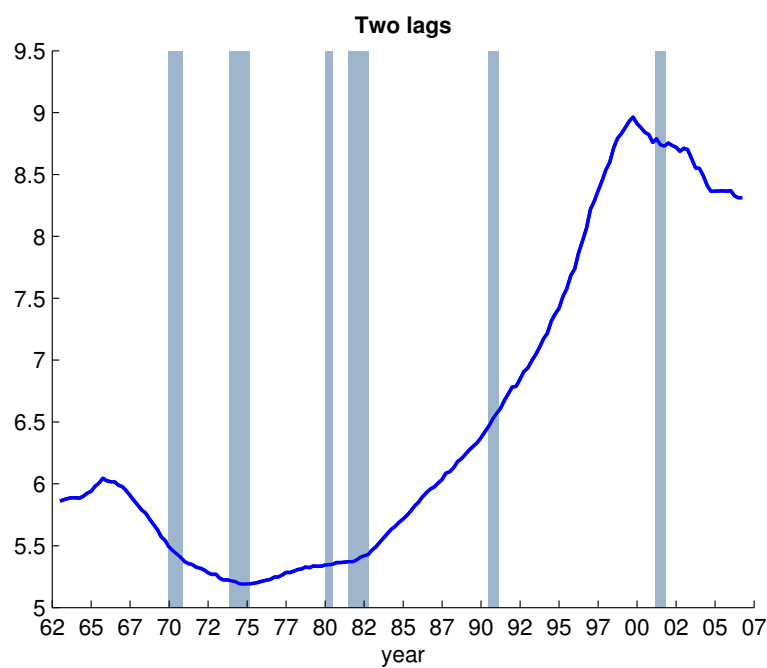

Figure 20: Robustness check, two lags; time-varying covariance between tax revenues and GDP conditional on non-policy shocks; the NBER recession periods are indicated by the shaded areas