DALLASFED Working Paper

The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply to Jentsch and Lunsford

Karel Mertens and Morten O. Ravn

Federal Reserve Bank of Dallas
Research Department
Working Paper 1805
https://doi.org/10.24149/wp1805

The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply to Jentsch and Lunsford*

KAREL MERTENS †
FRB Dallas, CEPR

MORTEN O. RAVN
University College London, CEPR, ESRC Centre for Macroeconomics

May 2018

Abstract: In this reply to a comment by Jentsch and Lunsford, we show that, when focusing on the relevant impulse responses, the evidence for economic and statistically significant macroeconomic effects of tax changes in Mertens and Ravn (2013) remains present for a range of asymptotically valid inference methods.

Keywords: Fiscal Policy, Structural Vector Autoregressions, Tax Shocks
JEL Classification: C32, E62, H24, H25, H31, H32

---

*We thank Kurt Lunsford for helpful discussions and for sharing code. We are also grateful to Ralf Brüggeman for providing the monthly version of the Kilian (2008) oil supply disruption series, to Alexander Chudik and José Montiel Olea for helpful discussions, and to Laton Russell for research assistance. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

†*Contact*: Federal Reserve Bank of Dallas, mertens.karel@gmail.com, tel: +(214) 922-6000

1 Introduction

Mertens and Ravn (2013) develop a methodology for estimating dynamic causal effects using proxies for shocks in structural vector autoregressive models (SVARs) and apply it to estimate the impact of personal and corporate income tax changes using narratively identified changes in taxes as proxies. To construct confidence intervals, we adopt a (multivariate) version of the recursive wild bootstrap, see Gonçalves and Kilian (2004), applied earlier in the literature for inference in SVARs, e.g. Kilian (2009). The main conclusion from the paper is that tax shocks have statistically significant effects on output and, depending on the type of tax, also on other key macro aggregates such as employment, spending on consumer durables, and investment.

In a recent contribution, Brüggeman, Jentsch, and Trenkler (2016) show that wild bootstraps are in general not asymptotically valid for inference about estimators that involve the covariance matrix of VAR innovations. While wild bootstraps correctly recover the asymptotic distribution of reduced-form impulse responses as the sample size increases, Brüggeman et al. (2016) show this is not generally the case for structural impulse responses. They present Monte Carlo evidence suggesting that wild bootstrap intervals for recursively-identified impulse responses may understate the true estimation uncertainty in finite samples. Jentsch and Lunsford (2018) point out that the results in Brüggeman et al. (2016) also apply to the wild bootstrap in Mertens and Ravn (2013). They propose a variant of the moving block bootstrap described in Brüggeman et al. (2016) as an alternative inference approach in proxy-identified SVARs. Based on the resulting 68% confidence intervals, Jentsch and Lunsford (2018) no longer find statistically significant effects. They conclude that "*... cuts to personal and corporate tax rates have no inferable effect on output, investment, employment, hours worked per worker, or the unemployment rate.* "

Based on the results in Brüggeman et al. (2016), we acknowledge that the wild bootstrap as applied in Mertens and Ravn (2013) is not generally asymptotically valid for proxy SVARs. We also view the moving block bootstrap as a potentially useful tool for inference in proxy SVARs. We do not agree, however, with the conclusion that there is no inferable effect of tax cuts on economic activity. We first explain why many of the intervals shown in Jentsch and Lunsford (2018) are not the relevant ones isolating the effects of changes in personal versus corporate income taxes. Next, we reconsider the empirical evidence on the impact of tax changes applying a number of alternative procedures for inference that are equally asymptotically valid. We show that this leads to the conclusion that tax shocks do have a significant impact on the economy as found by Mertens and Ravn (2013). More specifically, significance remains when we compute intervals using the Delta method or a parametric bootstrap described in Montiel Olea, Stock and Watson (2017), when we use the Jentsch and Lunsford (2018) intervals but a slightly different version of the proxies, or when we construct the percentile intervals exactly as in Brüggeman et al. (2016).

Asymptotic validity does not guarantee the reliability of an inference procedure in practice. In finite samples there is no a priori reason to prefer the Jentsch and Lunsford (2018) intervals over any of the available asymptotically valid alternatives. Any such preference must be motivated by Monte Carlo evidence using a data generating process that reasonably resembles that actual data used in a given application. We do not believe that Jentsch and Lunsford (2018) provide such evidence, nor do they compare the performance of their bootstrap with the other available approaches to inference in proxy SVARs. Moreover, the results in Mertens and Ravn (2013) are part of a much larger body of evidence for significant output effects of tax policy changes that does not rely on the use of wild bootstraps, or even proxy SVARs.[1]

A number of other applications of proxy SVARs have emerged in the literature, including to monetary policy shocks (Gertler and Karadi, 2015), uncertainty shocks (Carriero et al., 2015), oil shocks (Montiel Olea et al., 2017; Braun and Brüggeman, 2018), and credit supply shocks (Mian, Sufi and Verner, 2017). We present a brief comparison of the different inference approaches in some of these other applications, and we find that differences between confidence intervals tend to become more meaningful when the value of the Montiel Olea et al. (2017) test statistic for instrument relevance is relatively small. This suggests to us that instrument strength, rather than the issues raised by Jentsch and Lunsford (2018), may be the dominant concern for inference in most applications.

2 Inference in Proxy SVARs

We briefly revisit the proxy SVAR approach and provide a brief overview of existing methods for inference besides the Jentsch and Lunsford (2018) block bootstrap.[2]

Let $\mathbf{Y}_t$ be a $n \times 1$ vector of observables with a finite order VAR representation:

$$\mathbf{Y}_t = \sum_{j=1}^{p} \delta_j \mathbf{Y}_{t-j} + \mathbf{u}_t \ , \tag{1}$$

---

[1] See Mertens (2018) for a recent overview of US and international time series evidence.

[2] Because of its joint use of SVARs and instrumental variables techniques, Stock and Watson (2012) refer to the proxies as ‘external instruments’, while Stock and Watson (2018) have label the approach ‘SVAR-IV’.

where $\delta_j, j = 1, \ldots, p$ are $n \times n$ coefficient matrices and $\mathbf{u}_t$ is an $n \times 1$ vector of reduced-form innovations with covariance matrix $\Sigma_{\mathbf{uu}'}$. The key SVAR assumption is that the one-step-ahead forecast errors $\mathbf{u}_t$ (the innovations) are linear combinations of a vector of mutually orthogonal structural shocks ε_t:

$$\mathbf{u}_t = \mathbf{B}\varepsilon_t, \tag{2}$$

where ε_t is $n \times 1$, $\mathbb{E}[\varepsilon_t] = \mathbf{0}$, $\mathbb{E}[\varepsilon_t\varepsilon_t'] = \mathbf{I}_n$, $\mathbb{E}[\varepsilon_t\varepsilon_s'] = \mathbf{0}$ for $s \neq t$, and the $n \times n$ matrix $\mathbf{B}$ contains the contemporaneous causal effects of the structural shocks on the observables.

The coefficients in $\delta = [\delta_1, .., \delta_p]$ and the innovations $\mathbf{u}_t$ are easily obtained by least-squares, but more assumptions are needed to obtain any of the columns in $\mathbf{B}$ and obtain causal effects. Proxy SVARs arrive at identification by making use of available proxies/external instruments for the structural shocks of interest to (partially) identify the columns in $\mathbf{B}$. Suppose one is interested in identifying the first k columns of $\mathbf{B}$, corresponding to ε_{1t}, the first k shocks in ε_t. In addition, suppose k external instruments $\mathbf{m}_t$ are available that satisfy the following conditions:

$$\begin{aligned} &\text{(Relevance)} && \mathbb{E}[\mathbf{m}_t\varepsilon_{1t}'] = \Phi \,, \\ &\text{(Exogeneity)} && \mathbb{E}[\mathbf{m}_t\varepsilon_{2t}'] = \mathbf{0} \,, \end{aligned}$$

where Φ is an unknown $k \times k$ non-singular matrix, and ε_{2t}' contains the $n-k$ other structural shocks in ε_t. The first condition requires that the instruments $\mathbf{m}_t$ are relevant, i.e. that they are contemporaneously correlated with the true structural shocks ε_{1t}. However, it does not impose perfect correlation with the structural shocks of interest and therefore addresses the potential measurement errors related to instruments such as narratively identified shocks. The second condition requires that the instruments are exogenous. i.e. they are not contemporaneously correlated with the other structural shocks in ε_{2t}. Note, however, that the instruments $\mathbf{m}_t$ may still be correlated with lagged values of any of the structural shocks. Moreover, the instruments may be correlated with each of the shocks in ε_t, i.e. Φ need not be diagonal.

Mertens and Ravn (2013) show that the relevance and exogeneity conditions identify η, ζ and $\mathbf{S}_1\mathbf{S}_1'$ in

$$\mathbf{u}_{1t} = \eta\mathbf{u}_{2t} + \mathbf{S}_1\varepsilon_{1t} \,, \tag{3}$$

$$\mathbf{u}_{2t} = \zeta\mathbf{u}_{1t} + \mathbf{S}_2\varepsilon_{2t} \,, \tag{4}$$

where $\mathbf{u}_{1t}$ contains the first k forecast errors in $\mathbf{u}_t$, and $\mathbf{u}_{2t}$ contains the remaining $n-k$ errors. At this point, the only remaining object required to recover the first k columns of $\mathbf{B}$ is the $k \times k$ matrix $\mathbf{S}_1$. Because $\mathbf{S}_1\mathbf{S}_1'$ is identified by the proxies, this requires only $(k-1)k/2$ further restrictions. Most applications in the literature to date use external instruments that are assumed to be correlated with only a single structural shock, i.e. $k = 1$, in which case no further restrictions are necessary. The tax shock application in Mertens and Ravn (2013) is instead an example of $k = 2$ and the paper imposes the additional assumption that $\mathbf{S}_1$ is either upper or lower triangular.

For inference, Mertens and Ravn (2013) suggest to use a wild bootstrap to construct confidence bands for impulse response functions associated with personal and corporate tax shocks. The procedure involves (i) multiplying the reduced-form innovations and the proxies with random draws from the Rademacher distribution (-1 or 1 with equal probability), (ii) recursively constructing artificial samples of the observables using the estimated autoregressive parameters, (iii) obtaining the structural impulse responses using the bootstrapped proxies. Gonçalves and Kilian (2004) provide simulation evidence that wild bootstraps perform well in the presence of conditional heteroscedasticity in autoregressive models. The procedure seems a natural choice in the application to tax shocks because of the many zero observations in the proxies.

Montiel Olea, Stock and Watson (2017) is the first paper to develop theory for inference in SVAR models identified with external instruments, including the Delta method and a parametric bootstrap both of which are valid under strong-instrument asymptotics. Montiel Olea et al. (2017) also propose an inference approach that is asymptotically valid under weak-instrument asymptotics. The inference procedures in Montiel Olea et al. (2017), however, deal only with the case of a single external instrument, i.e. $k = 1$. Mertens and Montiel Olea (2018) apply the methods in Montiel Olea et al. (2017) to construct confidence intervals for impulse responses to marginal tax rate shocks, and also develop extensions of the Delta method and the parametric bootstrap in Montiel Olea et al. (2017) for the $k = 2$ case and a Newey and West (1987) residual covariance matrix. Unfortunately, weak-instrument robust intervals for the $k = 2$ case are currently not yet available.

While beyond our scope, we note that a number of recent studies develop Bayesian inference methods for proxy SVARs. Examples include Drautzburg (2016), Caldara and Herbst (2018), Miranda-Agrippino and Rey (2018), and Arias, Rubio-Ramirez and Waggoner (2018). Recent contributions by Antolin-Diaz and Rubio-Ramirez (2018) and Braun and Brüggeman (2018) also develop interesting extensions that combine narrative variables with sign restrictions.

3 Confidence Intervals for Mertens and Ravn (2013)

In the application to personal and corporate tax shocks, equation (3) can be written as:

$$\begin{bmatrix} u_t^{APITR} \\ u_t^{ACITR} \end{bmatrix} = \eta \mathbf{u}_{2t} + \mathbf{S}_1 \begin{bmatrix} \varepsilon_t^{APITR} \\ \varepsilon_t^{ACITR} \end{bmatrix}, \tag{5}$$

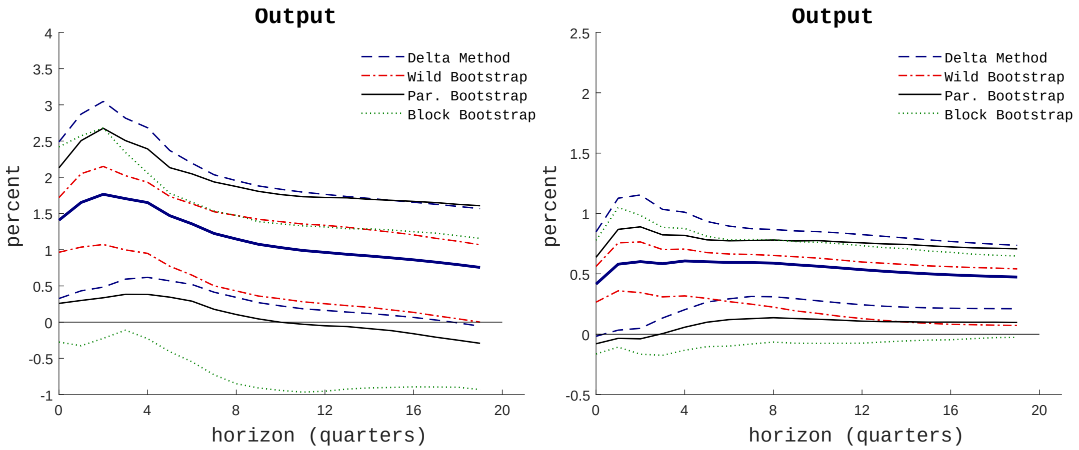
where APITR (ACITR) stands for the average personal (corporate) tax rate. Figures 2 and 3 in Mertens and Ravn (2013) show that imposing that $\mathbf{S}_1$ is either upper or lower triangular yields very similar results in the benchmark specification. Figures 1 and 2 in Jentsch and Lunsford (2018) instead show large differences in the confidence bands depending on the ordering of the tax rates. Specifically, the bands for the APITR (ACITR) shock are considerably more narrow when the APITR shock is ordered second (first).

We first explain why it is only the narrower bands in Jentsch and Lunsford (2018)'s Figures 1 and 2 that are relevant for judging the separate effects of changes in personal or corporate tax shocks. To make the interpretation of the different orderings clearer, we follow the discussion in Mertens and Montiel Olea (2018), and multiply both sides of (5) by the inverse of $\mathbf{S}_1$. Defining $\mathbf{C} \equiv \mathbf{S}_1^{-1}$ and $\gamma \equiv \mathbf{S}_1^{-1}\eta$, this yields:

$$\mathbf{C} \begin{bmatrix} u_t^{APITR} \\ u_t^{ACITR} \end{bmatrix} = \gamma \mathbf{u}_{2t} + \begin{bmatrix} \varepsilon_t^{APITR} \\ \varepsilon_t^{ACITR} \end{bmatrix}. \tag{6}$$

Consider first the zero restriction $C_{2,1} = 0$. This upper triangular restriction on $\mathbf{C}$ (or equivalently on $\mathbf{S}_1$) yields the response to an unanticipated counterfactual tax reform that, after controlling for current and lagged values of all endogenous variables with innovations $\mathbf{u}_{2t}$ as well as for lags of the tax rate variables, affect the APITR but has no impact on the ACITR. The associated impulse response is therefore due to an unexpected change the APITR, since any direct effect of the APITR shock on the ACITR in equation (6) is restricted to be zero on impact. The same upper triangular restriction $C_{2,1} = 0$ also identifies the response to an unexpected change in the ACITR, but allowing for a simultaneous exogenous change in the APITR. Consider next the zero restriction $C_{1,2} = 0$. This lower triangular restriction on $\mathbf{C}$ corresponds to a counterfactual tax reform that changes the ACITR but leaves the APITR unchanged, again after allowing for immediate feedback to the variables with innovations $\mathbf{u}_{2t}$. The lower triangular restriction $C_{1,2} = 0$ also identifies the response to a shock to the APITR, but allowing for an direct impact on the ACITR.

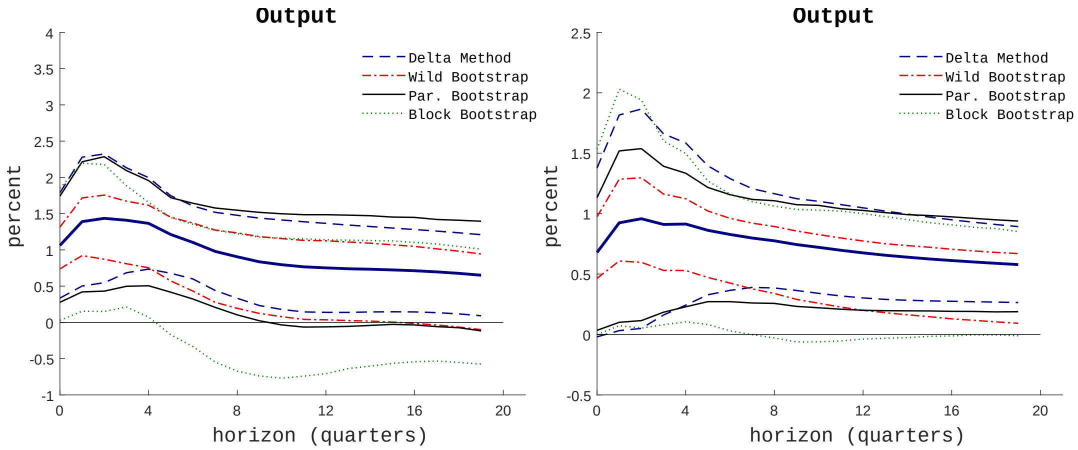
Different restrictions on $\mathbf{C}$ do not necessarily lead to meaningful differences in results in practice, but they generally do change the nature of the implied impulse response. Only the $C_{2,1} = 0$ restriction isolates the effects of an unexpected change in the APITR, while only the $C_{1,2} = 0$ restriction isolates the effects of an unexpected change in the ACITR. These are precisely the restrictions generating the narrower bands in Figures 1 and 2 of Jentsch and Lunsford (2018). They are also the restrictions imposed in Figures 9 and 10 of Mertens and Ravn (2013), showing the broader macroeconomic effects of personal and corporate tax rate shocks. The left columns in Figures 3, 4 and 5 of Jentsch and Lunsford (2018), in contrast, all show responses to an APITR change imposing that $C_{1,2} = 0$, which effectively are responses to some combination of changes in both tax rates. For these reasons, we will focus exclusively on the responses for the appropriate orderings/restrictions.



Notes: Bootstrapped intervals are based on 5000 replications.

Figure 1: Output response to APITR shock (Left) and ACITR shock (Right) with 68% standard percentile intervals.

Figure 1 shows the output responses to tax shocks for the benchmark specification in Mertens and Ravn (2013). We scale the tax shocks so that the average personal and corporate income tax rate, respectively, decline by one percentage point. The figure replicates the 68% Jentsch and Lunsford (2018) intervals and those generated by the wild bootstrap in Mertens and Ravn (2013). In addition, Figure 1 also shows new intervals based on the Delta method and the parametric bootstrap in Montiel Olea et al. (2017), each extended to the $k = 2$ case and a Newey and West (1987) HAC-robust residual covariance matrix as in Mertens and Montiel Olea (2018). The parametric bootstrap is based on draws from a joint normal distribution for all the model parameters using the estimated values and covariance matrix, see Montiel Olea et al. (2017). Consistent with Jentsch and Lunsford (2018), the wild bootstrap yields intervals that are clearly narrower than any of the alternative methods. Nonetheless, whereas the Jentsch and Lunsford (2018) intervals include zero at all horizons, this is not the case when we use the Delta method or the Montiel Olea et al. (2017) bootstrap. Either of these alternative procedures are asymptotically valid, and they all support the conclusion that tax shocks have significant impact on the economy.

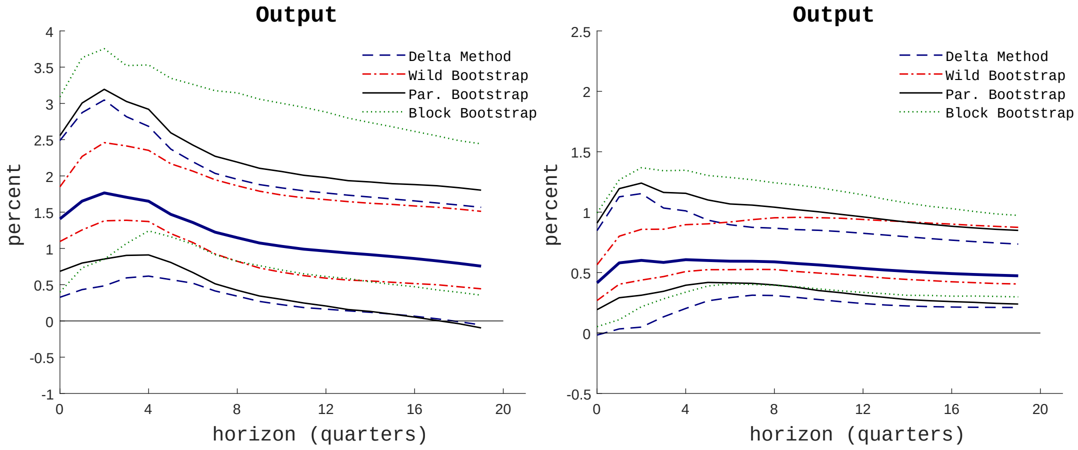


Notes: Bootstrapped intervals are based on 5000 replications.

Figure 2: Output response to APITR shock (Left) and ACITR shock (Right) with 68% standard percentile intervals and using the (uncentered) narrative shocks as proxies.

Our next illustration of the sensitivity of the claims made by Jentsch and Lunsford (2018) concerns a particular implementation detail. In Mertens and Ravn (2013), we constructed the proxies for the tax shocks by subtracting the mean from the nonzero observations of the narrative measures. The Jentsch and Lunsford (2018) bootstrap algorithm includes an additional centering of the non-censored proxies (see step 4 of their procedure) to ensure the bootstrap distribution has the same mean as the original proxies. The double centerings may however distort some of the informational content of the original data given the relatively small number of nonzero observations. Figure 2 repeats the analysis, but using the original narrative measures as the proxies, i.e. without removing the mean from the nonzero observations (which is not a requirement for the analysis). The resulting impulse responses remain very similar, with a slightly smaller (larger) output effect of an APITR (ACITR) decrease. Most of the confidence intervals also remain similar to those in Figure 1. The Jentsch and Lunsford (2018) intervals, however, again change qualitatively, and now in both cases exclude zero at short horizons.

A final experiment suggesting caution about the conclusions by Jentsch and Lunsford (2018) involves a difference in the inference procedure relative to Brüggeman et al. (2016).



Notes: Bootstrapped intervals are based on 5000 replications.

Figure 3: Output response to APITR shock (Left) and ACITR shock (Right) with 68% Hall (1992) percentile intervals.

All the bootstrap intervals reported so far are standard Efron and Tibshirani (1993) percentile intervals. The simulation evidence in Brüggeman et al. (2016) supporting the superior performance of the moving block bootstrap in finite samples is based on the percentile intervals proposed by Hall (1992). Figure 3 repeats the analysis of Figure 1, but now reports the Hall (1992) rather than the Efron and Tibshirani (1993) intervals. Asymptotically, both methods for constructing the intervals are equivalent, yet Figure 3 shows that the Hall (1992) intervals from the moving block bootstrap are very different from the Efron and Tibshirani (1993) intervals in Figure 1 and, in sharp contrast, are far away from zero. The other bootstrap intervals instead remain more similar to those in Figure 1.

We take away from this that the 68% bands for the relevant output responses obtained from available alternative procedures are indeed generally wider than those obtained from the wild bootstrap. At the same time, Figures 1, 2 and 3 provide a range of asymptotically valid alternatives producing 68% intervals that do not include zero. We also obtained analogous results for the effects on the other macroeconomic aggregates reported in Mertens and Ravn (2013). On this basis we therefore do not agree with Jentsch and Lunsford (2018)'s claim that the effects of tax shocks on output are not inferable. Finally, we note that the Jentsch and Lunsford (2018) intervals at the 90% or 95% levels become extremely wide, much more so than any of the other intervals. We conjecture that this is more likely a symptom of artificially low instrument relevance in large areas of the bootstrap distribution, in part because of the many zero observations in the tax proxies, than it is because the moving block bootstrap provides a better approximation of the true small sample distribution than the other inference approaches.

4 Confidence Intervals in Other Applications

In this final section, we explore the implications of adopting alternative inference approaches in other recent applications of the Proxy SVAR methodology. Each of these involve a single instrument and a single structural shock, such that $k = 1$ rather than $k = 2$ as in Mertens and Ravn (2013). This has the advantage that in these applications we can use available 'first-stage' statistics testing the relevance condition required for identification. Montiel Olea et al. (2017) derive the F-statistic for the null hypothesis that $\mathbb{E}[m_t \varepsilon_{1t}] \neq 0$ and show that its 'center' is analogous to the 'concentration' parameter in the linear IV model. This F-statistic provides an indication of possible weak-instrument concerns for inference, with the 5% critical value of 3.84, and the Stock and Yogo (2005) threshold value of 10 as useful reference points.

The first row in Figure 4 shows responses of GDP (left panel) and income reported to tax authorities (right panel) to a decrease in the average marginal tax rate, estimated by Mertens and Montiel Olea (2018). We consider the benchmark VAR estimated by Mertens and Montiel Olea (2018) which consists of 9 variables (real income per tax unit, log of one minus the average marginal tax rate, real output, unemployment, government spending, the change in federal debt, the inflation rate, real stock prices and the federal funds rate). The data are annual, the sample period is 1948-2012 and the VAR has two lags.

As before, the figures show 68% standard percentile intervals from the wild bootstrap in Mertens and Ravn (2013), the moving block bootstrap of Jentsch and Lunsford (2018), the Delta method and the parametric bootstrap in Montiel Olea et al. (2017). The proxy used for identification is a (weighted) average impact on statutory tax of a selection of historical US tax reforms. The F-statistic of 11.09 indicates a strong instrument. Comparing the intervals, the main observation is that the differences between the various 68% intervals are relatively minor, and certainly much less pronounced as in Figure 1 above.[3] Mertens and Montiel Olea (2018) show that at the 95% level the Jentsch and Lunsford (2018) become outliers and are much wider than the alternatives. Figure 5 shows the

---

[3] Mertens and Montiel Olea (2018) perform a similar evaluation of different confidence intervals in the appendix. The moving block bootstrap bands reported in Mertens and Montiel Olea (2018) uses the same algorithm as Jentsch and Lunsford (2018), but without the centering of the proxies in step 4 of their procedure. This yields 68% bands that are even closer to all the other bands.

block bootstrap distribution of the relevant first-stage F-statistic. More than 56% of the moving block bootstrap replications have an F-statistic smaller than 10, and 8% have a value smaller than 3.84. We suspect that weak instrument problems are potentially distorting the Jentsch and Lunsford (2018) intervals at higher significance levels, and that a similar phenomenon occurs in the Mertens and Ravn (2013) application above.

The next row in Figure 4 shows the response of the 1 year Treasury rate (left panel) and industrial production (right panel) to a monetary policy shock, estimated by Gertler and Karadi (2015). The proxy in this case consists of changes in interest rate futures shortly after FOMC announcements. As Gertler and Karadi (2015), we estimate a 12-lag monthly VAR model with 5 monthly macroeconomic and financial variables for the 1979:M7 to 2012:M6 sample.

The F-statistic is 9.15, which is marginally below the Stock and Yogo (2005) threshold. The 68% intervals around the interest rate response are very close to each other. The wild bootstrap intervals for the output response are more clearly narrower then the block bootstrap intervals. The latter in turn have somewhat smaller width than the Delta or the parametric bootstrap bands. Roughly 60% of the block bootstrap replications have an F-statistic smaller than 10, and 10% have a value smaller than 3.84.

The third row in Figure 4 shows the response of the VXO volatility index (left panel) and industrial production (right panel) to an uncertainty shock, as estimated by Carriero, Mumtaz, Theodoris and Theophilopoulou (2015). The proxy used is an indicator of geopolitical and other events constructed by Bloom (2009). Carriero, Mumtaz, Theodoris and Theophilopoulou (2015) estimate the 12-lag VAR model in Bloom (2009), which contains 7 monthly macroeconomic and financial variables for 1962:M7 to 2008:M6. The F-statistic is 22.30, which indicates a very strong instrument. The figure shows that there are no meaningful differences between the wild and block bootstrap intervals, and both are also very similar to the Delta and parametric bootstrap bands. Only a very small fraction of the block bootstrap replications has an F-statistic below 10.

The final row in Figure 4 shows the response of oil prices (left panel) and the Kilian (2009) index of global economic activity (right panel) to a negative oil supply shock. The proxy is a monthly version of the Kilian (2008) measure of oil supply shocks, as constructed by Braun and Brüggeman (2018). The variables in the 12-lag VAR are the growth rate of oil production, global real activity, and the real oil price over the 1973:M2 to 2007:12 sample. The F-statistic is 4.31, which exceeds the 5% critical value for rejecting instrument irrelevance, but is also considerably below the Stock and Yogo (2005) threshold. This value signals possible weak instrument problems for this application, see also Montiel Olea et al. (2017). There are considerable differences between the various confidence intervals. The wild bootstrap produces bands that are narrower than any of the others. The other bands are more similar in width, but meaningful differences remain across them. Virtually all of the block bootstrap replications have an F-statistic below 10.

We conclude from Figure 4 that the differences between the alternative 68% intervals are not necessarily large, and in some applications they are negligible. Moreover, the differences between intervals become more pronounced when the F-statistic for instrument relevance is lower. This statistic is not available for the $k = 2$ case in the Mertens and Ravn (2013) application, but the pattern across the applications in Figure 4 suggests that the differences between the intervals are related to the relevance of the tax proxies. While more research is needed on the performance of the various inference approaches, these findings suggest that instrument strength is in practice the important concern for inference.

5 Conclusion

We have considered the estimation of dynamic causal effects using proxy SVARs, and specifically the extent to which confidence intervals are sensitive to the choice of inference method. Our results show that the conclusions about the economic and statistical significance of the macroeconomic effects of tax changes in Mertens and Ravn (2013) remain broadly valid. We stress that proxy SVARs are a useful new tool for dynamic causal analysis by allowing researchers to combine the appealing nature of VAR analysis with new sources of identification. Research on inference in proxy SVARs is ongoing, and undoubtedly much progress will be made over the coming years in this important area of macroeconometrics.

References

Antolin-Diaz, Juan and Juan F. Rubio-Ramirez, 2018, “Narrative Sign Restrictions,” *American Economic Review*, forthcoming.

Arias, Jonas E., Juan F. Rubio-Ramirez and Daniel F. Waggoner, 2018, “Inference in Bayesian Proxy-SVARs,” manuscript.

Bloom, Nicholas, 2009, “The Impact of Uncertainty Shocks,” *Econometrica* 77(3), 623–685.

Robin Braun and Ralf Brüggeman, 2018, “Identification of SVAR Models by Combining Sign Restrictions With External Instruments,” manuscript, University of Konstanz.

Brüggeman, Ralf, Carsten Jentsch, and Carsten Trenkler, 2016, “Inference in VARs with Conditional Heteroscedasticity of Unknown Form,” *Journal of Econometrics* 191(1), 69–85.

Caldara, Dario and Edward Herbst, 2018, “Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVAR,” *American Economic Journal: Macroeconomics* forthcoming.

Carriero, Andrea, Haroon Mumtaz, Konstantinos Theodoris and Angeliki Theophilopoulou, 2015, “The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach,” *Journal of Money, Credit and Banking* 47(6), 1223–38.

Drautzburg, Thorsten, 2016, “A Narrative Approach to a Fiscal DSGE Model,” FRB Philadelphia Working Paper No. 16–11.

Efron, Bradley and R. J. Tibshirani, 1993, “An Introduction to the Bootstrap”, Chapman & Hall, New York.

Gertler, Mark and Peter Karadi, 2015, “ Monetary Policy Surprises, Credit Costs, and Economic Activity,” *American Economic Journal: Macroeconomics* 7(1), 44–76.

Gonçalves, Silvia and Lutz Kilian, 2004, “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form,” Journal of Econometrics 123(1), pp. 89–120.

Hamilton, James D., 2003, “What is an Oil Shock?” *Journal of Econometrics* 113 , 363–398.

Hall, Peter, 1992, “The Bootstrap and Edgeworth Expansion”, Springer, New York.

Jentsch, Carsten and Kurt Lunsford, 2016, “ Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States,” Federal Reserve Bank of Cleveland, Working Paper no. 16-19.

Jentsch, Carsten and Kurt Lunsford, 2018, “ The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment,” *American Economic Review* this issue.

Kilian, Lutz. 2008. “Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?” *The Review of Economics and Statistics* 90 (2), 216–240.

Kilian, Lutz. 2009. “Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market,” *American Economic Review* 99 (3), 1053–69.

Mertens, Karel, 2018, “ The Near Term Growth Impact of the Tax Cuts and Jobs Act,” Federal Reserve Bank of Dallas Working Paper 1803.

Mertens, Karel and José L. Montiel Olea, 2018, “ Marginal Tax Rates and Income: New Time Series Evidence,” *Quarterly Journal of Economics*, forthcoming.

Mertens, Karel and Morten O. Ravn, 2013, “ The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States,” *American Economic Review* 103(4), 1212–1247.

Mian, Atif, Amir Sufi and Emil Verner, 2017, “Household Debt and Business Cycles Worldwide,” *The Quarterly Journal of Economics* 132 (4), 1755-1817.

Miranda-Agrippino, Silvia and Hélène Rey, 2018, “US Monetary Policy and the Global Financial Cycle,” manuscript, London Business School.

Montiel Olea, José L., James H. Stock, and Mark W. Watson, 2017, “Inference in SVARs Identified with an External Instrument,” Columbia University Working Paper.

Newey, Whitney K. and Kenneth D. West, 1987, “ A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* 55, 703–708.

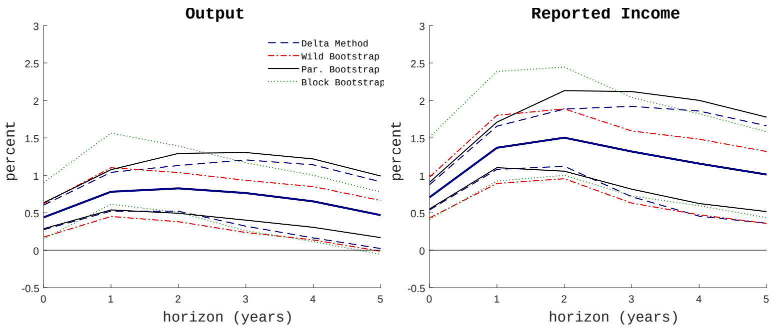
Stock, James H., and Mark W. Watson, 2008, “NBER Summer Institute Minicourse 2008: What’s New in Econometric Time Series, Lecture 7: Structural VARs,” at `http://www.nber.org/minicourse_2008.html`

Stock, James H., and Mark W. Watson, 2012, “Disentangling the Channels of the 2007-2009 Recession,” NBER Working paper no. 18094.

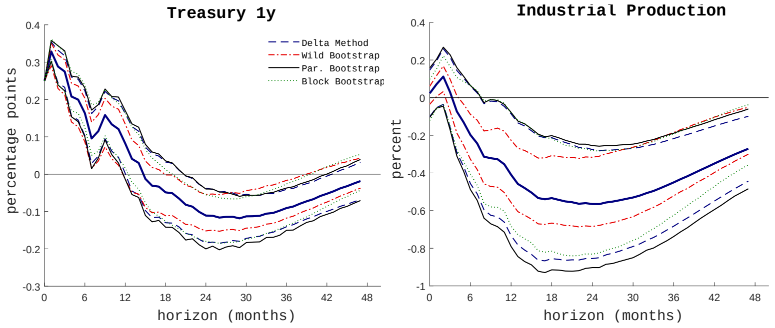
Stock, James H. and Mark W. Watson, 2018, “ Identification and Estimation of Dynamic Causal Effects in Macroeconomics,” *Economic Journal*, forthcoming.

Stock, James H., and Motohiro Yogo, 2005, “ IV Regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg ” D.W. Andrews and J.H. Stock, eds. (Cambridge, U.K.: Cambridge University Press, 2005).

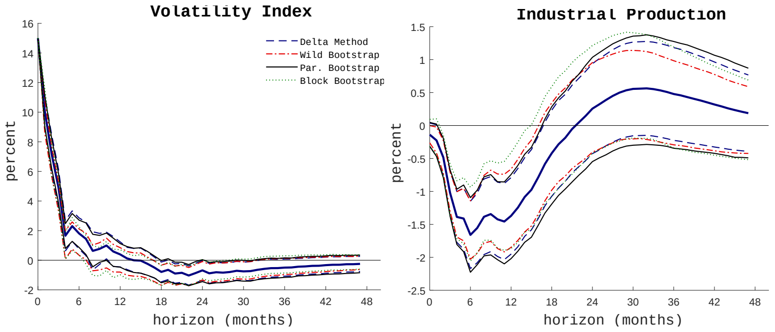
Mertens and Montiel Olea (2018) Marginal Tax Rate Shock (F-stat: 11.09)



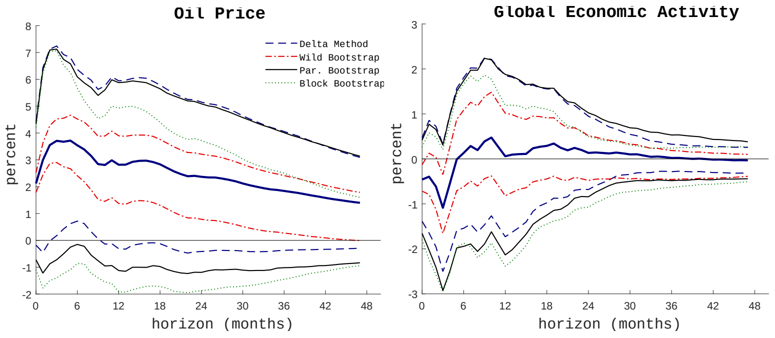
Gertler and Karadi (2015) Monetary Policy Shock (F-stat: 9.15)



Bloom (2009)-Carriero et al. (2015) Uncertainty Shock (F-stat: 22.30)



Kilian (2008, 2009) Oil Supply Shock (F-stat: 4.31)



Bootstrapped intervals are based on 5000 replications. The block length in the block bootstrap is the largest integer smaller than $5.03 \times T^{1/4}$, as suggested in Jentsch and Lunsford (2016).

Figure 4: Impulses and 68% Standard Percentile Intervals in Selected Applications of Proxy SVARs.

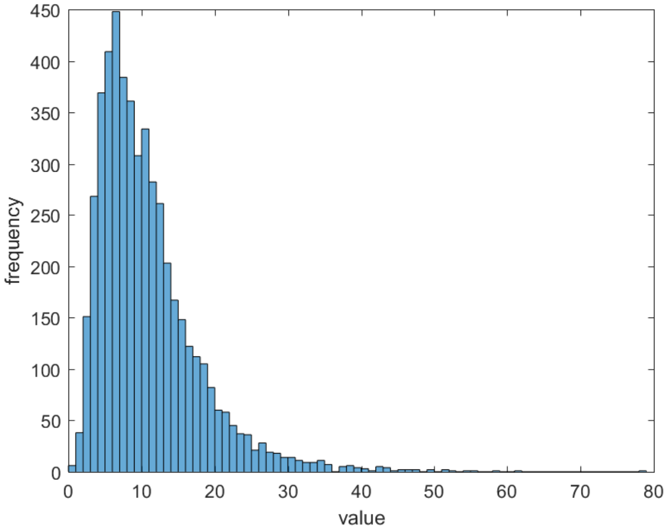


Figure 5: Mertens and Montiel Olea (2018): Block Bootstrap Distribution of the F-statistic