State-Dependent Macroeconomic Policy Effects: A Varying-Coefficient VAR*

Adrian Ochs † Christian Rörig †

April 5, 2022

Abstract

This article proposes a flexible framework to identify state-dependent effects of macroeconomic policies. In the literature, it is common to either estimate constant policy effects or introduce state-dependency in a parametric fashion. This, however, demands prior assumptions about the functional form. Our new method allows to identify state-dependent effects and possible interactions in a data-driven way. Specifically, we estimate heterogeneous policy effects using semi-parametric varying-coefficient models in an otherwise standard VAR structure. While keeping a parametric reduced form for interpretability and efficiency, we estimate the coefficients as functions of modifying macroeconomic variables, using random forests as the underlying non-parametric estimator. Simulation studies show that this method correctly identifies multiple states even for relatively small sample sizes. To further validate our method, we apply the semi-parametric framework to the historical data set by Ramey & Zubairy (2018) and offer a more granular perspective on the dependence of the fiscal policy efficacy on unemployment and interest rates.

Keywords: Varying-Coefficient Model, VAR, Random Forest, Heterogeneous Policy Effects, Fiscal Multiplier, State Dependent Policy Effects

JEL Codes: C32, E62

*Draft version; please do not distribute without the expressed permission of the authors

†University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom; e-mail: acro2@cam.ac.uk and cmr61@cam.ac.uk

1 Introduction

Endogeneity plagues the empirical study of policy effects. More often than not, any attempt to directly estimate the effect of the policy variable, x, on the response variable, y, will result in biased coefficient estimates since y might have given rise to the implementation of policy, x, but x affects y. Hence, a large part of the macroeconomic literature has focused on resolving this simultaneity issue by imposing short-run, long-run, or sign restrictions. More recently, constructing estimates of unanticipated policy shocks using them as external instruments. Particularly the use of external instruments has increased the confidence in coefficient estimates that no longer rely on sometimes hard to defend identification schemes (e.g. Wolf, 2020).

The compelling research on identification strategies to overcome endogenity concerns allows us to shift our attention to a hitherto often ignored source of estimation bias: model misspecification. With some exceptions, most empirical contributions on estimating macroeconomic policy effects assume a linear and constant effect structure. In contrast, many theoretical contributions highlight that effects could be non-linear due to asymmetric effects (i.e. an interest rate increase might not have the same effect as a decrease) and state-dependence (i.e. the economic environment as well as time). For instance, as Nakamura & Steinsson (2018) point out that monetary and fiscal policy effects may differ depending on the level of slack and openness of the economy. Consequently, even with a cleanly identified policy shock, estimating a linear model will yield biased estimates of the possibly non-linear response to policy.

To reduce the potential for model misspecification bias, we propose a flexible framework to identify state-dependent and asymmetric effects of macroeconomic policies. The approach uses a semi-parametric varying-coefficient estimator in an otherwise standard vector autoregression (VAR). Hence, we call this new approach a varying-coefficient VAR or VC-VAR. While keeping a parametric reduced form for interpretability and more efficient estimation, we estimate the (dynamic) policy effect non-parametrically by using random forests. Random forests allow for an entirely data-driven estimation of potentially heterogeneous policy effects, which can depend on the time or state of the economy. The proposed method could also be interpreted as a machine-learning augmented version of a threshold VAR (TVAR) model with the random forest predicting the splits and interactions of states.

The suggested approach to estimate VC-VARs can therefore be summarised in three steps. Firstly, determine the variables that may alter the VAR coefficient estimates. We call them moderators following the varying-coefficient literature. Secondly, define the VAR coefficients as functions of the set of moderator variables. Thirdly, estimate the VC-VAR using random forests.

In simulation studies, we show the advantages of the proposed semi-parametric estimation method. The VC-VAR:

1. Detects any state-dependence and interactions between different moderator variables in a dynamic system.
2. Captures non-linear and asymmetric effects, depending on the magnitude and the sign of the policy shock.
3. Performs well in identifying heterogeneous policy responses even for small sample sizes typical for the macroeconomic context. Thus, our approach offers a preferred bias-variance trade-off if the data generating process exhibits state-dependence.
4. Handles many moderator variables and detects the relevant moderator variables even when some moderators are correlated.

Furthermore, we apply the VC-VAR to the historic data-set by Ramey & Zubairy (2018) (RZ) to empirically validate the new method. RZ study the state-dependence of the fiscal multiplier for high or low unemployment and Zero Lower Bound (ZLB) periods. We are able to replicate their results by employing their pre-defined dummy variables for high or low unemployment and ZLB as moderator variables. Furthermore, we show that the newly proposed method yields again similar results when using the federal funds and unemployment rate directly as moderators. Hence, the VC-VAR method does not require the definition of dummy variables to detect state-dependence.

The empirical exercise also allows for several extensions of RZ's study. Firstly, using the federal funds rate and the unemployment rate as moderator variables offers a more granular perspective on the dependence of the fiscal multiplier on interest and unemployment rates and allows to analyse their interactions. This reveals that it is the unemployment rate that drives the state-dependence of fiscal policy effects and not the ZLB. We find that periods of ZLB mostly coincide with periods of high unemployment. Hence, not taking account of the interaction of unemployment and the ZLB would falsely indicate that fiscal spending was more effective in periods of ZLB. Secondly, using the instrument for the government spending shock as a moderator variable itself, we show that the asymmetric effect of fiscal shocks is small (i.e. positive or negative shocks have similar effects). Finally, we extend RZ's study by allowing for thirteen potential moderator variables from the FRED dataset. This reveals that the S&P 500 index and newly built housing units are important moderator variables for the estimation of state-dependent VC-VARs. These variables have not received much attention in previous theoretical or empirical research. The VC-VAR approach can deal with many potential moderators and therefore, allows for such exploratory analysis showing that fiscal spending appears to be more effective whenever the S&P 500 index and/or the construction of new housing units is growing.

Literature. This paper contributes to several strands of the literature. First, we provide a general method to estimate state-dependent policy effects. Recently, both theoretical (Rendahl, 2016; Canzoneri et al., 2016; Eichenbaum et al., 2018; Farhi & Werning, 2016) and empirical (Auerbach & Gorodnichenko, 2012, 2013; Fazzari et al., 2015; Ascari & Haber, 2019; Ramey & Zubairy, 2018; Paul, 2020) work has shed light on the importance of considering state-dependent and non-linear effects of macroeconomic policies. However, this has required to pre-specify any time- and state-dependence. More specifically, one has to construct dummies based on some discretionary thresholds and decide on the functional form of their interaction with the policy shock. This paper uses machine learning methods to automate this process and aims to shed more light on the high-dimensionality of policy effects. Using an underlying non-parametric estimator, we are agnostic about the functional form of any state-dependence.

Second, the new estimation method makes VARs more flexible, thereby reducing model-specific bias. Recent research on the differences of local projections (LP) and SVARs has described the choice between the two methodologies as a trade-off between variance and bias in applied research[1] (Li et al., 2021). While our approach is also concerned with misspecification bias, the focus is on a different source for this bias. Here, the bias stems from state-dependent coefficients. For Li et al. (2021), the model misspecification arises when the true data generating process (DGP) is a vector autoregressive moving average (VARMA) model, but the econometrician specified a VAR or when not selecting the correct lag length. LP's repeated estimation makes it more flexible for the latter misspecification but at the cost of increased variance. However, this does not hold for misspecifying a non-linear model as a linear one. Neither LP nor VARs can mitigate the bias from such misspecification. Nevertheless, the one-time estimation of the VAR model makes coefficient estimates with our method less volatile than LPs for small samples, which we usually observe in macroeconomics. Hence, we present our method in the context of VARs. This might make VARs more attractive when suspecting state-dependence and choosing between the two methodologies, LPs and VARs. Similarly, Ramey & Zubairy (2018) have argued that LPs allow for more straightforward inclusions of non-linearities. Our new estimation method makes this process simple and automatic for the VAR case, focusing the choice between LPs and VARs back on statistical properties rather than the simplicity of implementation.

Third, we follow a recent strand of research exploring how to utilize advances from the machine learning literature for macroeconomic analysis (Duarte, 2018; Joseph, 2019; Goulet Coulombe, 2020; Gu et al., 2020). Our paper is closest to Goulet Coulombe (2020) who also proposes a varying-coefficient model in macroeconomics. However, the main difference is that Goulet Coulombe (2020) aims to improve macroeconomic predictions. We are more interested in identifying potential state-dependence in macroeconomic policies using a high-dimensional semi-parametric VAR framework. Hastie & Tibshirani (1993) first introduced varying coefficient models. These

[1] When estimated in population with infinite lag length, both methods produce the same IRFs (Plagborg-Møller & Wolf, 2019)

models are linear in their regressors, but their coefficients are estimated by generalized additive models to allow them to vary as a function of other factors. Wang & Hastie (2014), for instance, use regression trees as the base learner to identify varying effects. The advantage of using regression trees over other non-parametric techniques, such as kernel-based estimates, is that they can handle many potentially mixed-type variables. Zhou & Hooker (2019) propose to use boosted decision trees as the non-parametric effect modifiers and provide proof of asymptotic consistency of the estimator.

Outlook. The rest of the paper is structured as follows. Section 2 gives an overview of the varying-coefficient model and motivates the use of random forests as the non-parametric coefficient estimator. Furthermore, section 2 explains how to estimate high-dimensional impulse response functions using the varying-coefficient model. In section 3, we conduct several simulation studies showcasing the performance of the method, while section 4 applies the method to the recent paper by Ramey & Zubairy (2018), estimating state-dependent fiscal multipliers. Finally, section 5 concludes.

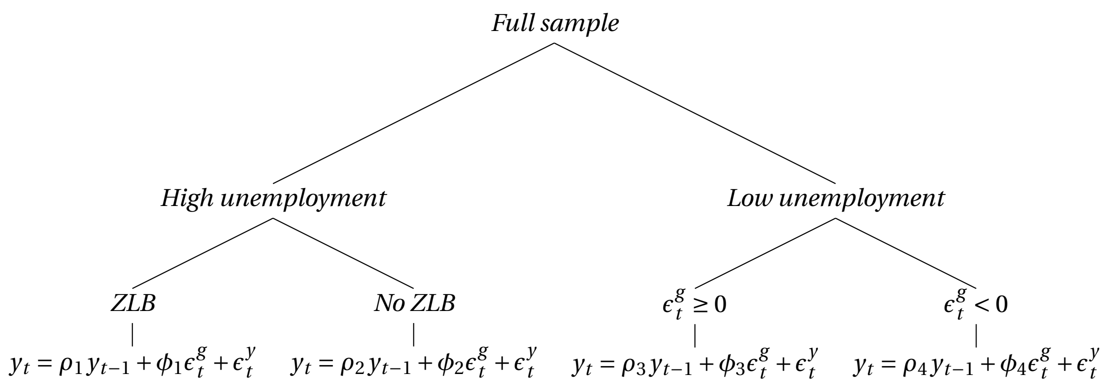
2 Methodology

This section first motivates the use of trees as the underlying non-parametric estimator for capturing regime shifts in the economy. Next, we introduce a varying-coefficient model using trees as base learners and provide computational details on how to estimate state-dependent coefficients non-parametrically. Furthermore, we show how this estimator can be used in the context of structural vector autoregressions (SVAR) to obtain high-dimensional impulse-response functions.

2.1 Macroeconomic state-dependencies as trees

Random forests as non-parametric estimators have increased in popularity as they allow for complex non-linearities and asymmetric effects that can handle high-dimensional data and are, compared to other machine learning algorithms, robust as they demand little to no tuning.

Figure 1: Example Tree



Note: Illustrative tree capturing state-dependent policy effects. The set of potential moderators is $L = \{Unemployment_{t-1}, ZLB_{t-1}, \epsilon_{t-1}^g\}$, where ϵ^g is the government spending shock.

Moreover, random forests offer a natural way to represent the macroeconomic state-space and state-dependence in macroeconomic time-series (see also Goulet Coulombe (2020)). Anticipating our empirical application, Figure 1 shows an illustrative tree capturing potential state-dependence of fiscal policy efficacy as in Ramey & Zubairy (2018). Prior to estimation, a researcher studying fiscal policy might assume that the effect of government spending shocks denoted by ϵ_t^g could depend on the level of unemployment as suggested by Rendahl (2016) and found by Ramey & Zubairy (2018). This is depicted in the first split of Figure 1 into the "high unemployment" and "low unemployment" branch. Furthermore, for "high unemployment", it might matter whether the economy is already close to the ZLB or not. If this were the case, the tree-based estimator would split the "high unemployment" branch into two further branches differentiating between "ZLB" and "no ZLB". Additionally, the output response might be asymmetric in the "low unemployment" state, depending on the sign of the policy shock. This would create a further split of the "low unemployment" branch into "$\epsilon_t^g \geq 0$" and "$\epsilon_t^g \leq 0$". This step-by-step illustration shows how macroeconomic state-dependencies can be depicted by trees.

For parametric estimation approaches, the assumptions depicted in figure 1 pose a high degree of model uncertainty. To avoid model specification bias, one has to correctly pre-define the different states and cut-off levels, as well as their interactions. Using tree-based learners in our semi-parametric approach, we reduce the problem of model uncertainty to the choice of which effect modifiers to include (e.g. unemployment and ZLB), while the random forest estimates the underlying splits as illustrated in figure 1.

2.2 Varying Coefficient Model

This subsection introduces a varying coefficient model which employs tree-based learners to augment the estimation of a parametric baseline specification in search of potential state-dependence.[2]

Consider the following simple model structure:

$$
\begin{aligned}
y_t &= \beta_t \epsilon_t + \boldsymbol{x}_t' \gamma_t + u_t, \\
\beta_t &= f(\Omega_t),
\end{aligned}
$$

where y_t is some macroeconomic outcome variable of interest, ϵ_t some exogenous policy shock, $\boldsymbol{x}_t$ a vector of controls, and β_t is a high-dimensional coefficient, which is estimated non-parametrically as a function of the state space of the economy, Ω_t. We will assume throughout that, ϵ_t, $\boldsymbol{x}_t$ and Ω_t, are exogenous. In this setup, the relationship between the outcome, y_t and some policy shock, ϵ_t may vary depending on the state of the economy captured by Ω_t. In principle, Ω_t can contain many modifying variables and it may overlap with other exogenous regressors in the low-dimensional linear regression.

[2]Another advantage of using trees as base learners over non-parametric kernel smoothing or spline-based methods is that they can handle high-dimensional data of mixed types.

Tree-based varying coefficient model. In the definition of the tree-based estimator, we follow Wang & Hastie (2014)[3]. The tree-based method aims to approximate $\beta(\Omega_t)$ by a piece-wise constant function. The idea is to partition the state space Ω_t following a certain optimization criterion and to approximate $\beta(\Omega_t)$ in each partition by a constant vector. Let $C_m{}_{m=1}^{M}$ denote a partition of the state space $\mathbb{R}^q$ satisfying $C_m \cap C_{m'} = \emptyset$ for any $m \neq m'$, and $\cup_{m=1}^{M} C_m = \mathbb{R}^q$, where M denotes the number of partitions. Then, the tree-based varying coefficient estimator can be written as

$$y_t = \sum_{m=1}^{M} \beta_m I_{(\Omega_t \in C_m)} \epsilon_t + \boldsymbol{x}_t' \gamma_t + u_t, \tag{1}$$

where $I_{(\bullet)}$ denotes the indicator function with $I_{(c)} = 1$ if event c is true and zero otherwise. The implied heterogeneous policy effect is thus,

$$\beta(\Omega_t) = \sum_{m=1}^{M} \beta_m I_{(\Omega_t \in C_m)}.$$

Consequently, a constant OLS estimate of β can be interpreted as the average policy effect over all partitions

$$\beta_{OLS} = \sum_{m=1}^{M} \frac{n_m}{N} \beta_m I_{(\Omega_t \in C_m)}.$$

where n_m and N denote the observations in partition m and the overall sample size, respectively. The set partitions C_m are also referred to as terminal nodes or leaf nodes of the tree, describing the various estimated states of the policy effect. However, the number of nodes M, the particular partitions $\{C_m\}_{m=1}^{M}$ as well as the values for β_m are unknown and require simultaneous estimation. The number of partitions is usually tuned using some criteria evaluating the out-of-sample fit. The latter is crucial in order to avoid over-fitting. For instance, picture the case of a tree that is partitioned entirely based on the in-sample fit. The best and, in fact perfect fit would be reached by having N partitions, each representing a single observation. There are several methods to avoid over-fitting, with cross-validation being the most common one.

[3] Wang & Hastie (2014) use a tree-based estimator as a base learner for boosting, while we will bag multiple tree-based estimates to obtain a more robust forest-based estimate.

During cross-validation, the sample is randomly split into a training and validation set, with the first being used for estimating the model and the latter for testing the model's performance out-of-sample. We use k-fold cross-validation where the sample is split into k folds such that k-1 folds are the training set and one fold is the test set. This is repeated k times for each fold.

Having decided on the number of partitions, which is usually done automatically by most implementations, the remaining problem is to find the optimal partition sets and coefficient estimates corresponding to those splits. Given the parametric nature of the governing function, we minimize according to the least squares criterion

$$\left(\hat{C}_m, \hat{\beta}_m\right) = \underset{(C_m, \beta_m)}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - \sum_{m=1}^{M} \beta_m I_{(\Omega_t \in C_m)} \epsilon_t - \boldsymbol{x}_t' \gamma_t \right)^2 \tag{2}$$

.

In the above specification, the estimation of β_m is nested in that of the partitions. Hence, within each partition, β_m is simply the least squares estimator on the corresponding sub-sample. Therefore, minimizing equation (2) boils down to finding the optimal set of partitions $\{C_m\}_{m=1}^{M}$ such that

$$\hat{C}_m = \underset{C_m}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{m=1}^{M} \left(y_t - \beta_m(C_m) I_{(\Omega_t \in C_m)} \epsilon_t - \boldsymbol{x}_t' \gamma_t \right)^2 \tag{3}$$

.

Forest-based varying coefficient model. The forest-based estimator represents a collection of many tree-based estimates following the random forest ensemble algorithm by Breiman (2001). The idea is simply to reduce the variance of a single tree's estimate by averaging over multiple trees. More specifically, k trees are estimated based on k independent identically distributed random samples of the data set. With each tree casting a prediction for β_m, the random forest isdictor is formed by taking the average over all trees. This procedure is commonly referred to as bootstrap averaging or bagging as introduced in Breiman (1996). Additionally, Breiman (2001) proposes to use a random subset of state space variables in Ω_t, which improves the accuracy by minimizing the correlation between state space variables while at the same time keeping the strength of their signal. Breiman (2001) shows that random forest are more robust and accurate compared to single trees. Given their preferred characteristics, we will use forest-based estimates throughout this paper.

The underlying optimization problem remains the same as in equation (3) for a single tree as a base learner with only slight changes to the tree producing algorithm now using random subsets of the state space at every split. The forest-based estimate of the varying coefficient is then simply the average over all these trees:

$$\beta_{RF}(\Omega_t) = \frac{1}{K}\sum_{k=1}^{K}\sum_{m_k=1}^{M_k}\beta_{m_k} I_{(\Omega_t \in C_{m_k})}.$$

Computational details. We rely (for now) on the implementation of a coefficient-wise varying coefficient regression model by Buergin & Ritschard (2017) readily available in their R-package `vcrpart`. While this section will give a short overview of how the estimation algorithm works, we refer for further details to the original paper. Algorithm 1 provides a formal summary of how to estimate a varying coefficient model using a random forest as its underlying non-parametric estimator.

The Random Forest-VCM function in Algorithm 1 is a wrapper function growing numerous trees based on bootstrapped samples of the original data. Coefficient predictions are then obtained by averaging over all tree predictions leading to reduced variance and hence more robust estimates.

The individual trees are estimated in the Randomized Tree-VCM function in Algorithm 1. The main optimization objective is to minimize the negative log likelihood of the linear parametric function. Given some minimal node size, i.e. the number of observations which are represented by each terminal node of the tree, and a minimum log-likelihood improvement required for each split, the algorithm searches for the optimal partition over the value space of the modifying variables for each coefficient at each node. This is done until no further split satisfies

Algorithm 1 Estimation of varying coefficient model using a random forest following Buergin & Ritschard (2017)

Parameters: T number of trees in forest, e.g., T = 100
N_0 minimum node size, e.g., $N_0 = 30$
$D_{\min}$ minimum $-2\cdot$ log-likelihood reduction, e.g., $D_{\min} = 2$
$mtry$ number of combinations of coefficients, nodes, and moderators to randomly sample as split candidates, e.g., $mtry = 5$

function RANDOM FOREST-VCM(S, Ω)
  $H \leftarrow \emptyset$ ▷ Initialize Forest
  for trees in $t = 1$ to T **do**
    $S_t \leftarrow$ A bootstrap sample from the dataset S
    $h_t \leftarrow$ Randomized Tree-VCM(S_t, Ω)
    $H \leftarrow H \cup h_t$
  end for
  return H
end function ▷ Coefficient predictions by averaging over all trees in H

function RANDOMIZED TREE-VCM(S, Ω)
  Initialize $\mathscr{C}_{k1} \leftarrow \Omega_{k1} \times \ldots \times \Omega_{kL}$ and $M_k \leftarrow 1$ for all partitions $k = 1, \ldots, K$.

  repeat
    Compute $\hat{l}^{\hat{\mathscr{M}}} = \max_{\beta,\gamma} l^{\hat{\mathscr{M}}}(\beta, \gamma)$ of the current model using the latest partitions

$$\widehat{\mathscr{M}} : y_i = \mathbf{x}_i' \boldsymbol{\gamma} + \sum_{k=1}^{K} \sum_{m=1}^{M_k} I_{(\Omega_{ik} \in \mathscr{C}_{km})} x_{ik} \beta_{km}$$

    using all observations $i \in S$.
    Randomly sample combinations of coefficients, nodes and moderators.
    for partitions $k = 1$ to K^{mtry} **do**
      for nodes $m = 1$ to M_k^{mtry} and randomized subset of moderators $\tilde{l} = 1$ to $\tilde{L}^{mtry}$ **do**
        for all unique candidate split Δ_{kmlj}, in $\{\omega_{kli} : \Omega_{ik} \in \mathscr{C}_{km}\}$ that divides $\mathscr{C}_{km}$ into two nodes $\{\mathscr{C}_{kmlj1}, \mathscr{C}_{kmlj2}\}$ and satisfies $\min_s \sum_i I_{(\Omega_{ik} \in \mathscr{C}_{kmljs})} \geq N_0$ **do**
          Using only the observations $\{i : \Omega_{ik} \in \mathscr{C}_{km}\}$ of the node $\mathscr{C}_{km}$, compute $\hat{\ell}\widehat{\mathscr{M}}_{kmlj} = \max_{\{\beta_1,\beta_2\}} \ell\widehat{\mathscr{M}_{kml}}(\beta_1, \beta_2)$ of the approximate search model

$$\widehat{\mathscr{M}}_{kmlj} : y_i^{(s)} = \hat{y}_i + \sum_{s=1}^{2} I_{(\Omega_{ik} \in \mathscr{C}_{kmljs})} x_{ik} \beta_s$$

          and compute the training error reduction $D_{kmlj} = -2\hat{\ell}\widehat{\mathscr{M}} + 2\hat{\ell}^{\widehat{\mathscr{M}}_{kml}}_{\{i:\Omega_{ik} \in \mathscr{C}_{km}\}}$, where $\hat{\ell}^{\widehat{\mathscr{M}}}_{\{i:\Omega_{ik} \in \mathscr{C}_{km}\}}$ is the subtotal of $\hat{\ell}\widehat{\mathscr{M}}$ for the observations of $\mathscr{C}_{km}$.
        end for
      end for
    end for
    Split node $\mathscr{C}_{k'm'}$ by $\Delta_{k'm'l'j'}$ where $D_{k'm'l'j'} = \max D_{kmlj}$ and increase $M_{k'} \leftarrow M_{k'} + 1$.
  until no candidate split satisfies N_0 or $D_{k'm'l'j'} < D_{\min}$
  return Return Tree-VCM
end function

the minimum number of observations in each terminal node, or the log-likelihood improvement is smaller than the minimum requested.

Similar to bootstrapping, growing individual trees with subsampling of combinations of coefficients, nodes, and moderators as split candidates reduces the variance. The subsampling decorrelates the individual trees since they have different split candidates. The decorrelation of the individual trees makes the averaging over all trees more effective in reducing the variance. Nevertheless, the number of combinations as split candidates is a tuning parameter and in some cases, no subsampling might yield the best out of sample fit. We find that this is mostly the case when we have few coefficients and moderators.

2.3 High-dimensional impulse response functions

This section shows how to utilize the semi-parametric estimator outlined above to estimate state-dependent effects in a vector autoregression (VAR) framework and subsequently how to obtain state-dependent IRFs (for a brief summary of the step-by-step estimation of state-dependent IRFs, see Algorithm 2 below.).

The key difference to the usual way of computing SVARs is that we impose the identifying restrictions directly on the structural equations rather than the reduced form covariance matrix. We first show that these approaches are equivalent and then clarify why direct estimation is beneficial when estimating state-dependent IRFs.

To illustrate our approach, assume a bivariate VAR(1) model and that there exists an internal instrument, x_t, such that the system can be ordered recursively and is just identified. Let us further assume that the internal instrument, x_t has a state dependent contemporaneous effect on the outcome variable, y_t. The structural model takes the following form.

$$\begin{aligned}\mathbf{A}Y_t &= \mathbf{B}Y_{t-1} + \epsilon_t \\ \begin{bmatrix} a_{11} & 0 \\ a_{21}(\Omega_t) & a_{22} \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^x \\ \epsilon_t^y \end{bmatrix}\end{aligned} \tag{4}$$

with the mean and variance of the structural errors normalized to zero and one, respectively and no correlation between the errors such that $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[\epsilon_t \epsilon_t'] = I_n$, where I_n is the identity matrix. The state dependence of the effect of the instrument, x_t on y_t is indicated by writing a_{21} as a function of the state, Ω_t. Let the state be defined as $\Omega = \{s_t, r_t\}$, which implies that the coefficient a_{21} might vary according to different values for s_t and r_t. Note that the state could include many more moderator variables that may overlap with other exogenous regressors.

Furthermore, we assume $\mathbf{A}$ to be a non-singular matrix enabling us to obtain the reduced form (5) by inverting $\mathbf{A}$:

$$\begin{aligned}Y_t &= \mathbf{F}Y_{t-1} + u_t \\ \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} f_{11} & f_{12} \\ f_{21}(\Omega_t) & f_{22}(\Omega_t) \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix}\end{aligned} \tag{5}$$

where $\mathbf{A}^{-1} \equiv \mathbf{Q}$, $\mathbf{F} = \mathbf{Q} * \mathbf{B}$ and $u_t = \mathbf{Q}\epsilon_t$.

Due to the state dependent effect, $a_{21}(\Omega_t)$ of the internal instrument on the outcome variable, using a linear estimator would misspecify the system and hence, lead to biased IRFs. This is the case for both, the structural and the reduced form, as a_{21}, f_{21} and f_{22} are all a function of the state, Ω_t. Hence, whenever state dependence is present, the usual SVAR identification procedure yields biased estimates.

If we are interested in the effect of a structural shock of x_t on y_t, we can also directly estimate the ratio of a_{21} and a_{22}. This is, given the restrictions, we can rewrite system (4) as

$$x_t = \frac{b_{11}}{a_{11}} x_{t-1} + \frac{b_{12}}{a_{11}} y_{t-1} + \frac{1}{a_{11}} \epsilon_t^x, \tag{6}$$

where we define $\frac{1}{a_{11}}\epsilon_t^x \equiv u_t^x$ and

$$\begin{aligned} y_t &= -\frac{a_{21}(\Omega_t)}{a_{22}}x_t + \frac{b_{21}}{a_{22}}x_{t-1} + \frac{b_{22}}{a_{22}}y_{t-1} + \frac{1}{a_{22}}\epsilon_t^y \\ &= -\frac{a_{21}(\Omega_t)}{a_{22}}u_t^x + \left(\frac{b_{21}}{a_{22}} - \frac{a_{21}(\Omega_t)}{a_{22}}\frac{b_{11}}{a_{11}}\right)x_{t-1} + \left(\frac{b_{22}}{a_{22}} - \frac{a_{21}(\Omega_t)}{a_{22}}\frac{b_{12}}{a_{11}}\right)y_{t-1} + \frac{1}{a_{22}}\epsilon_t^y. \end{aligned} \tag{7}$$

Hence, this allows us to directly estimate the impact effect, $-\frac{a_{21}(\Omega_t)}{a_{22}}$, using a varying coefficient model for y_t,

$$y_t = \beta_1(\Omega_t)x_t + \beta_2(\Omega_t)x_{t-1} + \beta_3(\Omega_t)y_{t-1} + u_t^y. \tag{8}$$

Note that β_1 in equation (8) is the relative impact effect on y_t of a structural shock that increases x_t by one unit. This is, rather than identifying q_{21}, direct estimation gives the scaled effect $\frac{q_{21}}{q_{11}}$, which can be interpreted as the impact effect of a shock that increases x_t by one unit.[4] Mostly, this is the unit of interest for policy. Nevertheless, to obtain absolute impact effects (i.e. a one standard deviation shock), one can use the standard deviation of the residuals from estimating equation (6), which is equivalent to q_{11} and multiply it with β_1 to obtain q_{21}. Note that the standard deviation of the structural shocks is assumed to not change over the different estimated regimes. Hence, the scaling can be done with the residuals from equation (6) and (8) rather than obtaining them for every state.

The reason for using the direct estimation approach now becomes apparent. Instead of estimating a fully parameterized version of equation (8), we can apply the semi-parametric estimation method outlined above to obtain state-dependent estimates of β_1 in one step.[5] Using the more conventional approach and imposing the restrictions on the covariance matrix of the reduced form errors would necessitate estimating the reduced form for every regime and obtain the covariance matrix for every estimation. This would lead to efficiency losses and potentially unstable results when the amount of observations for a regime is small.

[4]To see this more clearly, note that $\begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22}}\begin{bmatrix} a_{22} & 0 \\ -a_{21} & a_{11} \end{bmatrix} = \begin{bmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{bmatrix}$.

[5]Paul (2020) includes instruments for the structural shock of interest as an exogenous variable in a VAR to more easily estimate state-dependent effects. Our direct estimation approach is also valid when there is no external instrument available.

It is straightforward to produce IRFs in a simple model where only one coefficient varies between two values. For the example above, assume that a_{21} takes on a different value for $s_t > s^*$ and $s_t <= s^*$. Then, we can obtain values for the impact effect for β_1 and similarly for f_{21} and f_{22} when $s_t > s^*$ and $s_t <= s^*$ and compute the two IRFs.

However, we usually do not know a priori which variables are state-dependent and whether there are interactions within the variables that determine the state dependence. For this reason, in practice, we estimate (5) fully flexibly to get the state dependent dynamic matrix, F, as:

$$Y_t = \mathbf{F}(\Omega_t)Y_{t-1} + u_t \tag{9}$$

and similarly the structural equation (8) to obtain the relative impact, $\beta_1(\Omega_t)$:

$$y_t = \beta_1(\Omega_t)x_t + \beta_2(\Omega_t)x_{t-1} + \beta_3(\Omega_t)y_{t-1} + u_t$$

Estimating the SVAR model fully flexibly complicates the computation of IRFs. Instead of displaying all possible IRFs under all combinations of the varying coefficients, we propose to predict the IRFs for different values of the state variables. That is, after estimating the model, one can define several states of interest, e.g.: $\Omega^* = \{s_t = s^*, r_t = r^*\}$, $\Omega^{**} = \{s_t = s^{**}, r_t = r^{**}\}$ and $\Omega^{***} = \{s_t = s^{***}, r_t = r^{***}\}$ to then compute the IRFs for each of these states. The state-dependent IRF more generally then becomes:

$$\begin{aligned}\widehat{IRF}_{t+h}(\Delta, \Omega_t) = \mathbb{E}\left[y^*_{t+h} \mid \epsilon_{t+\tau} = \begin{cases} \Delta & \text{if } \tau = 0 \\ 0 \text{ if } \tau \in (1,h) \end{cases} ; \Omega_t\right] \\ - \mathbb{E}\left[y_{t+h} \mid \epsilon_{t+\tau} = 0, \forall \tau \in (0,h); \Omega_t\right]\end{aligned} \tag{10}$$

where Δ is the size of the shock. IRFs are conditional on a specific state, i.e. specific values of the state variables, and we do not assume any regime switching.

We adapt the tree-based estimation to VARs rather than LPs since compared to a dummy approach that explicitly imposes the regime shift on the model, our flexible estimation technique might estimate different regimes at different horizons when using LPs. Additionally, even when the same regimes are found for every horizon, the point estimates might be more erratic due to the repeated estimation procedure and less efficient since we lose observations by iterating over the horizons. Appendix A shows that the confidence intervals become wider when producing state-dependent IRFs with LPs.

Algorithm 2 Estimating heterogeneous IRFs

Step 1 Decide on state variables in set $\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$.
Step 2 Estimate reduced form equations semi-parametrically using the set Ω as coefficient modifiers.
Step 3 Estimate structural equations semi-parametrically using the set Ω as coefficient modifiers.
Step 4 From estimated model in **2** predict lag coefficients for specific state values $\{\omega_1 = \omega_1^*, \omega_2 = \omega_2^*, ..., \omega_k = \omega_k^*\} \in \Omega$ and combine to $F(\omega^*)$.
Step 5 From estimated model in **3** predict contemporaneous impact coefficients for state values $\omega \in \Omega$ and combine to $Q(\omega)$.
Step 6 Get state-dependent IRFs by: $\widehat{IRF}_{t+h}(\Delta, \omega) = F^h(\omega)Q(\omega)\epsilon$, where ϵ is the vector of structural shocks.

---

3 Simulation studies

In this section, we showcase the performance of the VC-VAR in identifying heterogeneous policy responses in the data. We consider several different simulation environments from a simple case of one state-dependent coefficient and two states to multiple varying coefficients and various states. For the purpose of illustrating that the VC-VAR approach offers a more favourable bias-variance trade-off compared to an approach using local projections, we included simulation results for the latter in Appendix A (see Figures 13 - 16).

Exogenous structural break. Let's first consider the following dynamic system with an exogenous change in the effect of the policy instrument.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.3 & t < T/2 \\ -0.1 & t \geq T/2 \end{cases}
\end{aligned}
$$

where i_t is some policy instrument, e.g. the nominal interest rate, and y_t is some outcome variable of interest, e.g. output growth. The policy effect is piece-wise constant and there is a structural break amid the simulated sample. We simulate the system for $T = 500$ and estimate the policy effect using the VC-VAR without any priors on the number of structural breaks, their location or the values in the respective partitions. Since the focus of this exercise is to identify potential state- or time-dependence in the policy response, we assume to know the actual exclusion restrictions when estimating the system of equations, not to introduce any endogeneity bias. This is without loss of generality, given that any estimator would be biased if the exogeneity assumption is violated.

The resulting estimates for β_t are compared to OLS estimates and their distributions are plotted in Figure 2a. As expected, the OLS estimates are only able to capture the average policy effect[6]. The semi-parametric estimates, however, are able to distinguish between the different states of β_t clearly and are very close to the true values.

Figure 3a shows the corresponding impulse response estimates for a shock in the policy instrument. While the parametric OLS estimate exhibits smaller confidence bounds, it is clearly biased as it only captures the average response missing the structural break. The median state-dependent response of the semi-parametric estimator, on the other hand, is very close to the true impulse responses. In fact, when plotting the partial dependency[7] of the varying coefficient β_t, we can see that it found the exact structural break in the middle of the sample.

Endogenous change. In the next example we consider the policy effect being moderated by previous values of the outcome variable, y. While before we simulated a clear structural break, in this case the states can quickly switch endogenously. The piece-wise definition of β_t allows introducing all kinds of non-linearities and asymmetries in the policy response for the forest-based estimator to uncover.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.3 & y_{t-1} < 0 \\ -0.1 & y_{t-1} \geq 0 \end{cases}
\end{aligned}
$$

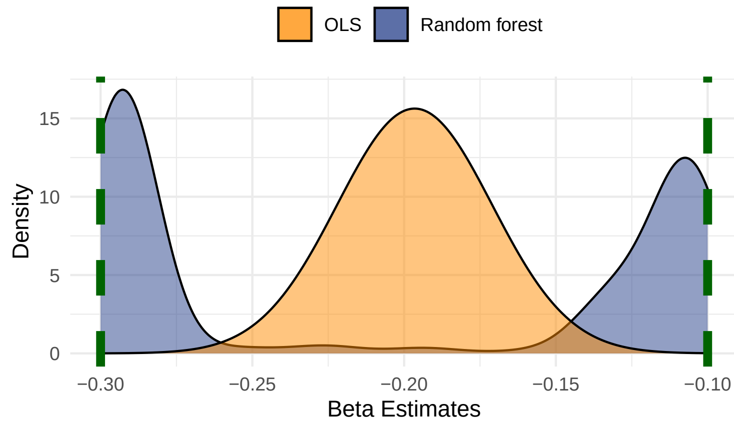
Figure 2b plots the resulting distributions of the estimates for β_t. Again, the forest-based estimator clearly identifies two states, while the OLS estimate solely identifies the average policy effect. The partial dependency plot in figure 3b shows that the semi-parametric estimator correctly identifies the split between positive and negative lagged values of y_t. Consequently, the estimated state-dependent impulse responses follow the actual ones closely.

[6]Clearly, if the policy effect is in fact constant, OLS is the best linear unbiased estimator (BLUE).

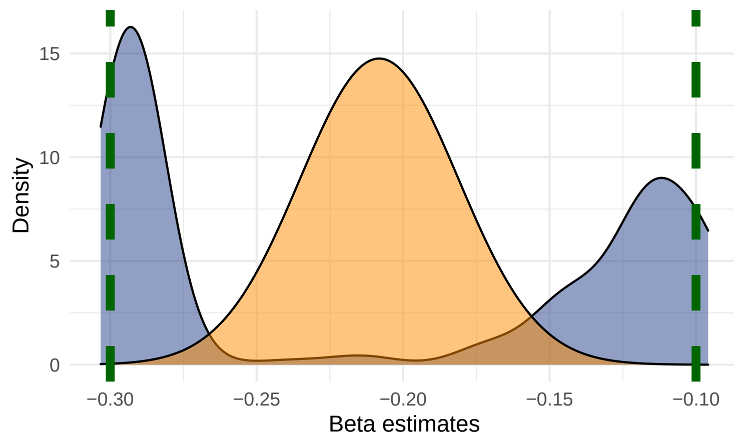
[7]The partial dependency is computed by predicting the coefficient for the domain of the modifying variable – here, time.

Figure 2: Coefficient Estimate Distributions - OLS and Random Forest

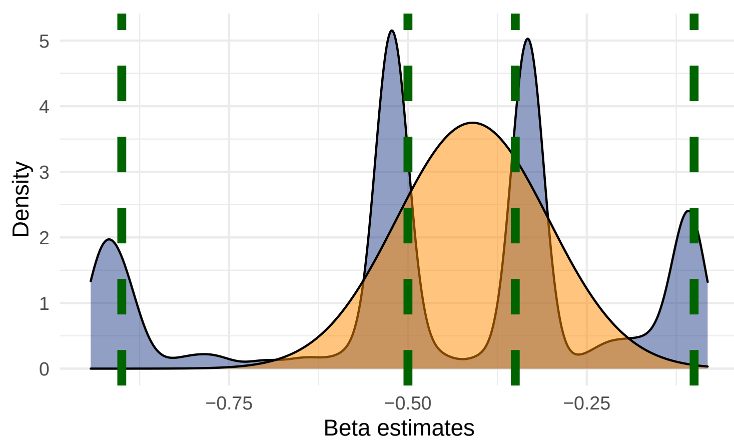
(a) Exogenous structural break



(b) Endogenous change



(c) Endogenous and exogenous change



Note: This figures compares the distribution of the estimated policy effects using OLS (orange) and Random Forest (blue) in different simulation environments. The vertical, dashed, green lines indicate the true coefficient values. Simulations are conducted with T = 500.

Exogenous and endogenous change. Next, we consider both, an exogenous structural break as well as an endogenous change in the policy effect over time.

$$i_t = 0.8 i_{t-1} + 0.1 y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2)$$
$$y_t = 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2)$$
$$\beta_t = \begin{cases} -0.1 & y_{t-1} < 0, t < T/2 \\ -0.35 & y_{t-1} \geq 0, t < T/2 \\ -0.5 & y_{t-1} < 0, t \geq T/2 \\ -0.9 & y_{t-1} \geq 0, t \geq T/2 \end{cases}$$

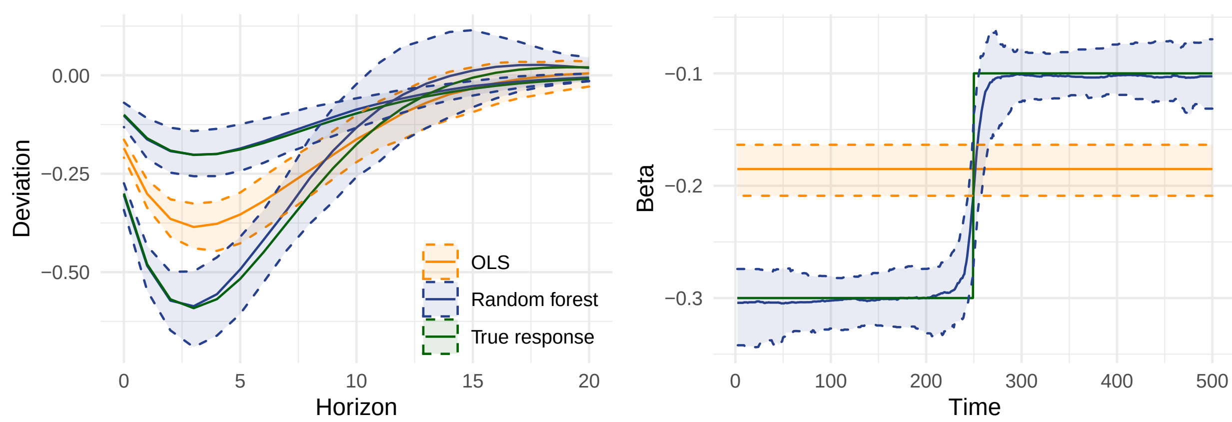
As we can see in figure 2c, the semi-parametric approach has no problem in identifying the four different states of β_t and gives a much more granular picture of the true policy effect. Figure 3c documents how closely the semi-parametrically estimated impulse response functions track the actual ones for all four states. In contrast, the OLS estimate only informs about the average response over the entire sample. The surrogate tree[8] illustrates how well the underlying random forest identifies the actual splits and state-dependent values of β_t.

Impact and dynamic change. So far, we have only considered cases in which only the impact coefficient, β_t, varied across states. However, it might also be that the policy rule is state-dependent, affecting the propagation of the shock. Let's consider the following dynamic system, with state-dependence in the impact effect of the policy and a varying policy parameter, γ_t.

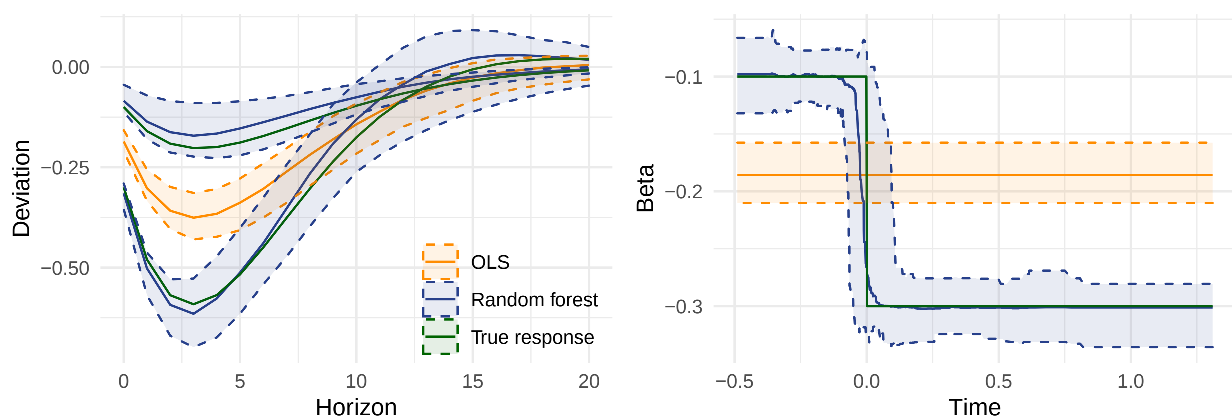
[8] Surrogate trees aim to illustrate the resulting splits by random forests with the latter being a collection of many trees. Obtaining a surrogate tree is done by fitting a single tree to the predictions of the random forest.

Figure 3: IRFs and Surrogate Trees - OLS and Random Forest

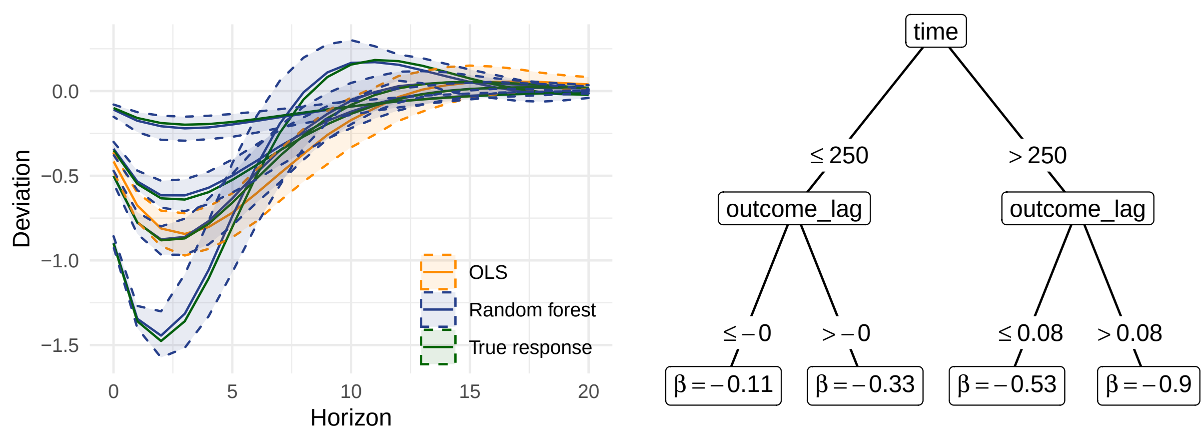
(a) Exogenous structural break



(b) Endogenous change



(c) Endogenous and exogenous change



Note: This figures compares the IRFs using OLS (orange) and Random Forest (blue) in different simulation scenarios. The green lines depict the true response. Simulations are conducted with T = 500.

$$
\begin{aligned}
i_t &= 0.8 i_{t-1} + \gamma_t y_{t-1} + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 0.5^2) \\
y_t &= 0.8 y_{t-1} + \beta_t i_t + \epsilon_t^y, \quad \epsilon_t^i \sim N(0, 0.25^2) \\
\beta_t &= \begin{cases} -0.1 & y_{t-1} < 0, t < T/2 \\ -0.35 & y_{t-1} \geq 0, t < T/2 \\ -0.5 & y_{t-1} < 0, t \geq T/2 \\ -0.9 & y_{t-1} \geq 0, t \geq T/2 \end{cases}, \quad \gamma_t = \begin{cases} 0.3 & y_{t-1} < 0 \\ 0.1 & y_{t-1} \geq 0 \end{cases}
\end{aligned}
$$

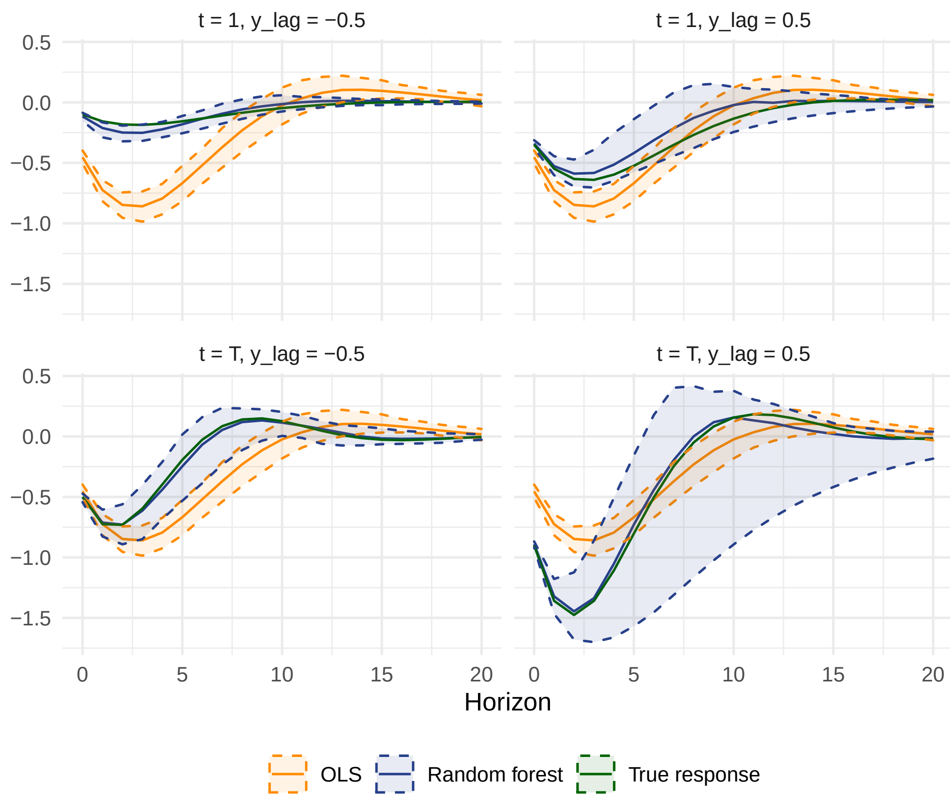
.

In general, with more lags and endogenous variables in the VAR structure, it becomes harder to justify why only single coefficients should be state-dependent. Hence, we will usually estimate more than one coefficient of the system semi-parametrically. This shifts the focus from considering state-dependent splits of a single coefficient as we did above to predicting the whole system of coefficients for specific state values.

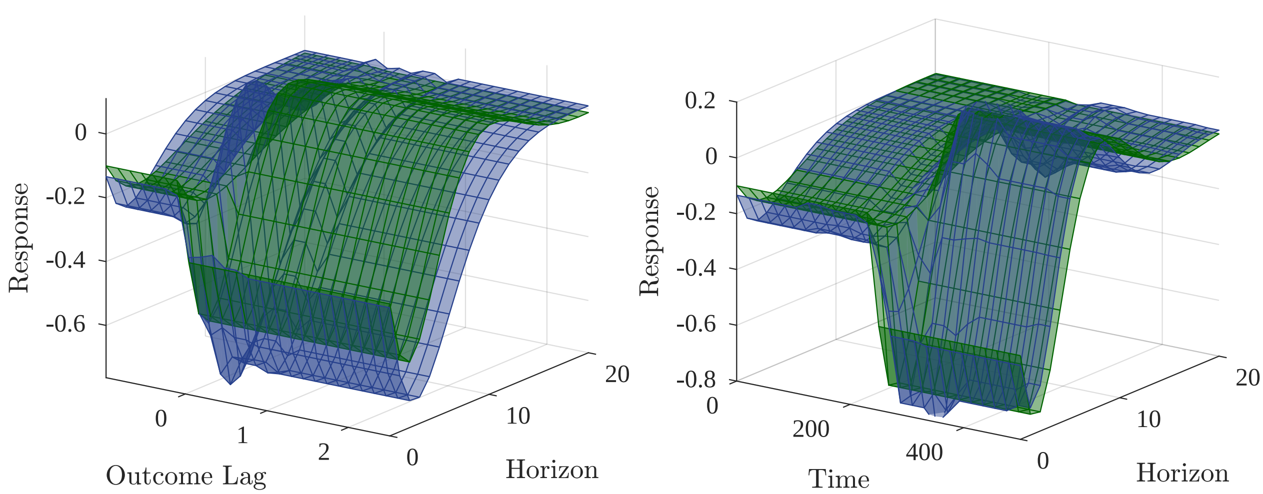
Figure 4a plots estimates of the impulse response after a shock to the policy instrument for specific states. While the OLS estimates only offer an aggregated view, the semi-parametric estimates capture state-dependence well. Instead of selecting specific state values to predict the system and subsequently the heterogeneous impulse responses, another way of presenting the functional form of the state-dependence is to compute partial dependency IRFs. Partial dependency IRFs are computed by varying one state over its domain and fixing all other states at their median values and predicting the impulse response. This leads to IRFs with the state of interest as an additional dimension as depicted in Figure 4b. We can see that the estimated high-dimensional impulse response in blue is very close to the true response in green and the split points are correctly identified.

Figure 4: IRFs and Partial Dependency Plots

(a) Impulse responses for different cases



(b) Partial dependency impulse responses



Note: This figures compares the IRFs and their partial dependency using OLS (orange) and Random Forest (blue) for impact and dynamic changes. The green lines depict the true response. Simulations are conducted with T = 500.

Many potential moderators. The previous simulation examples used the relevant moderators for the coefficient estimates to test whether our proposed estimator finds the correct splits. However, in practice, we might not know in advance which macroeconomic variables actually impact the policy response. In this example we feed several moderators to our estimator, of which only some are relevant. Furthermore, we allow for some correlation between the relevant and non-relevant moderators. Hence, this subsection shows that the VC-VAR approach can find the relevant moderators among many other, potentially correlated but non-relevant, moderators. We show this using a so-called variable importance measure (VIM) for random forests. Finding the relevant moderators among many candidates is vital in practice when researchers are unsure which moderators might impact coefficient estimates.

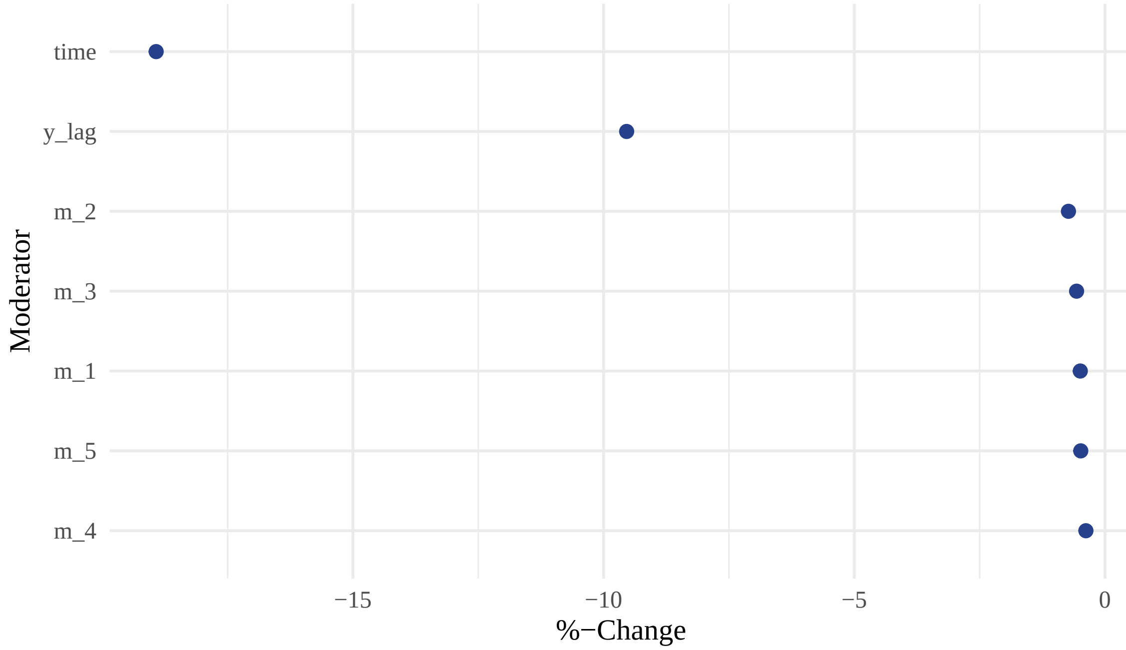
The varying coefficient model to be simulated and estimated is the same as for the exogenous and endogenous change above. Additionally, we include two moderator variables, m_1 and m_2, that are correlated with the relevant moderator variable, y_{t-1}, with a correlation coefficient of 0.7 and 0.3, respectively. Furthermore, we draw three non-relevant moderators, m_3, m_4, and m_5, from normal distributions. We include these variables to test whether the random forest can choose the relevant moderators among correlated and noise moderators.

To evaluate, which moderators are important for the estimation, we use a VIM. One could also eyeball the partial-dependency IRFs to gauge the relevance of each moderator. However, this would become infeasible with an increasing number of moderators and the VIM provides a numeric measure.

We calculate the VIM in three steps: (1) Estimate the varying coefficient model using all potential moderators. (2) Randomize the observations within one moderator and obtain the predicted values. (3) Obtain the out-of-sample mean squared error by subtracting the full model's predicted values from the model's predicted values with one moderator's observations randomized.

The idea behind this procedure is the following. If a moderator is important for the prediction, then feeding the observations of that moderator in a randomized order into the model

Figure 5: Variable Importance Measure Simulation



Note: This figure depicts the variable importance measure showing the percentage decrease in the mean squared error from the model with one randomized moderator to the full model. The simulation was conducted with 1000 observations and the Random Forest was estimated with 100 trees (increasing the number of trees did not change the results much).

would yield a worse prediction. Similarly, if a moderator is not relevant for prediction, even if we randomized its observations, that should not impact prediction. Hence, whenever the mean squared error decreases a lot when using the full model, we assume the moderator to be important.

Figure 5 presents the variable importance measures from our simulated model. Moderators $time$ and y_lag are correctly identified as the relevant moderators. For both, randomizing the order of their observations increases the means squared prediction error substantially, while the irrelevant variables have no additional predictive power. Thus, this simulation shows that the estimator can correctly distinguish the relevant moderators from other non-relevant ones.

Recall that moderators m_1 and m_2 are correlated with y_lag. Despite this correlation, the estimator only selects y_lag and *time* as important moderators and correctly classes m_1, and m_2 as non-relevant. Hence, even if non-relevant moderators are correlated with a relevant one, the VIM yields the correct classification.

4 Empirical application

This section illustrates the proposed methodology in an applied setting replicating and extending the study by Ramey & Zubairy (2018). The replication exercise reveals that our method yields similar results to conventional IRF estimation techniques. Furthermore, we show that the new methodology allows us to study state-dependence in more detail, extending the study by RZ.

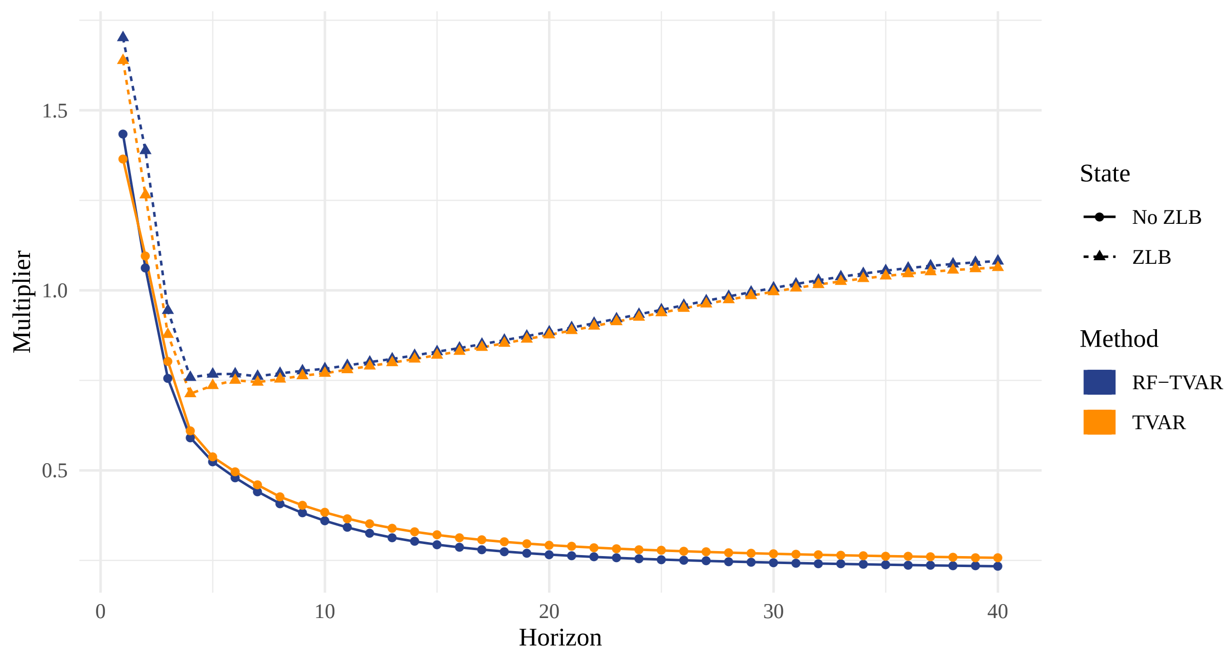
RZ study the effect of government spending shocks on output to investigate the government spending multiplier in low and high employment regimes as well as at the ZLB. To do so, they derive a historical time series of government spending news shocks and use them in local projections as well as TVAR analyses as an instrument for government spending. We focus our comparison on the TVARs since they are more similar to our setting. [9]

To study state-dependence of government spending effects, RZ construct a dummy for slack taking the value of zero whenever unemployment is below 6.5% and one in case it is equal or higher. Similarly, they define a dummy for periods in which the economy was close to the ZLB.

The definition of dummies and thereby pre-definition of states is restrictive. First, deciding on the threshold is usually either an ad-hoc assumption or quite the opposite and the result of trying a battery of different thresholds. Both approaches come with a high degree of model uncertainty. Second, dummies are binary in nature allowing only for two states. Of course, one could include multiple dummies to account for multiple states but then one would have to justify the choice of several thresholds. Our method can be seen as more flexible allowing us to estimate any underlying state-dependence of the VAR non-parametrically as well as implicitly estimating interactions between the states in one estimation step. Studying the interaction of

[9] Note that RZ use an LP-IV approach in their main analysis. Nevertheless, LP-IV can be shown to be equivalent to ordering the instrument first in an SVAR (Plagborg-Møller & Wolf (2019)). To obtain the same relative IRFs that the LP-IV approach yields, however, we have to divide by the impact impulse response on government spending. To compute the multiplier, the division by the impact effect on government spending is not necessary since we are interested in the ratio of the two IRFs of government spending and output. Thus, the division by the impact impulse would cancel out.

Figure 6: ZLB Dependent Multiplier with Different Methods



Note: This figure compares the estimated government multipliers for a ZLB and non-ZLB state. The multipliers are estimated using a TVAR (orange lines) and a Random Forest (blue lines) with the ZLB dummy directly fed in as a moderator.

states might be of interest to the researcher since often the states in questions are correlated. For example, it is plausible that high unemployment states are correlated with periods close to the ZLB.

4.1 Sanity check: Using Dummies as Modifier Variables

In a first benchmark exercise, we compare the RZ results to estimates obtained from our VC-VAR model when directly feeding in the dummy as the sole effect modifier. Given the pre-defined dummy, the tree-based learners can only either split in line with the dummy or not at all in case the coefficient is state-independent. Hence, by just including a dummy, the semi-parametric estimator offers the same amount of restricted flexibility as a TVAR. Consequently, feeding in the dummy should yield equivalent results to what RZ find.

Figure 6 depicts estimates of the fiscal multiplier in normal times and close to the ZLB. The orange lines correspond to the estimates found by RZ using a TVAR, while the blue lines refer to estimates from the semi-parametric VAR. It can be seen that the state-dependent multipliers converge to very similar values. Some small initial differences are expected since RZ compute the TVARs separately for each state[10], while the semi-parametric VAR makes use of the entire sample.

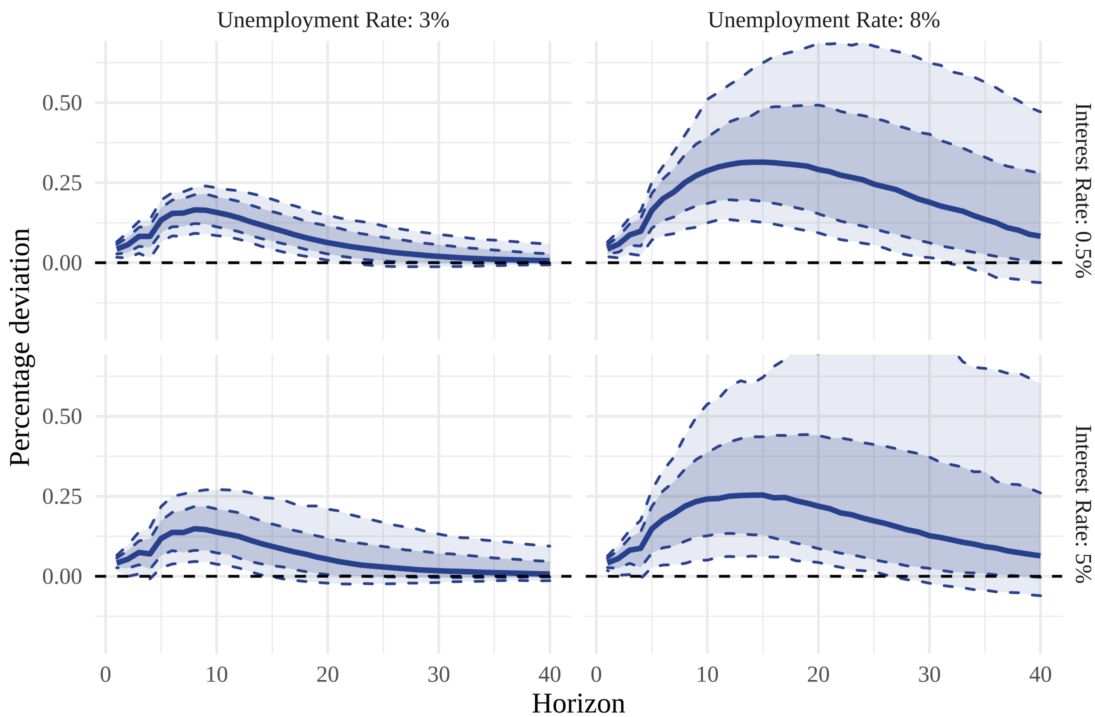
4.2 Semi-Parametric Estimation

In a next step we aim to illustrate the flexibility of the semi-parametric approach. First, in contrast to RZ we do not pre-define any states using dummy variables but directly feed in the continuous variables of the unemployment rate and the interest rate leaving it up to the VC-VAR to find appropriate splits. Second, while RZ estimate separate models looking for potential state-dependence of fiscal policy in regimes of high unemployment and times close to the ZLB, we estimate this simultaneously allowing for potential interaction between regimes. Moreover, we allow all coefficients in the structural VAR estimation to vary across different states of the unemployment rate, the interest rate, the world war dummy and their potential interactions. The result is an SVAR model estimate which can be predicted for different states. Hence, in order to compute IRFs, we simply predict the entire system of coefficients for a given combination of states. The underlying assumption of these high-dimensional IRFs is that there is no regime switching and the IRF is conditional on a specific state.

Visualizing high-dimensional IRFs is not as straightforward as for one-dimensional impulse responses. In the case of a low-dimensional state space, one can plot a battery of IRFs for combinations of specific values of the effect modifiers. Figure 7, for instance, shows the estimated impulse response functions for two different levels of the unemployment rate and the interest rate when the world war dummy is zero. The dark shaded areas depict the 68% and the lightly shaded areas the 90% confidence interval, while the solid line denotes the point estimate. The estimates suggest that fiscal policy is more effective in times of high unemployment. This is in

[10] According to the replication STATA file `tvar.do` by Ramey & Zubairy (2018).

Figure 7: IRF of Output

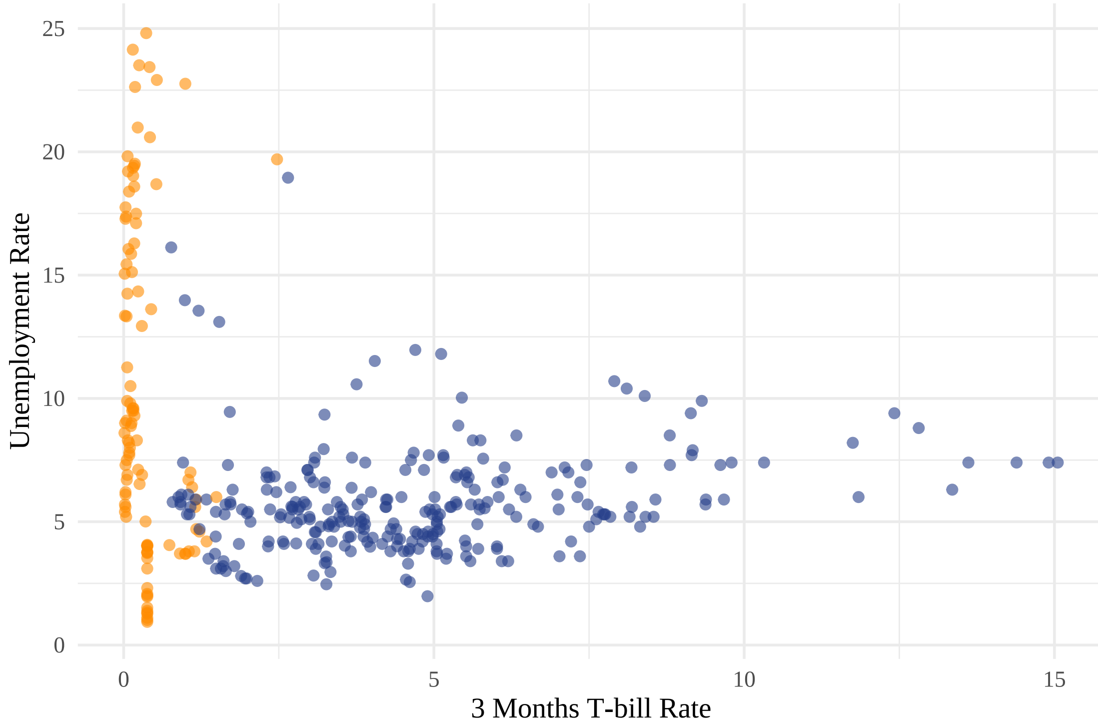


Note: This figure shows the IRFs of output following a fiscal policy shock. The IRFs are plotted for different combinations of low and high unemployment under a ZLB or no ZLB. The shaded areas depict 68% and 90% confidence intervals obtained with wild-Bootstrap of 500 iterations.

line with the findings by RZ and the theoretical contribution by Rendahl (2016).

However, when controlling for unemployment, the efficacy of fiscal policy is not higher close to the zero lower bound as RZ find. The difference might be due to the definition of the zero lower bound dummy in Ramey & Zubairy (2018). RZ define periods of continuous low interest rates and low inflation as binded by the zero lower bound without choosing a specific interest rate threshold. Furthermore, their zero lower bound dummy spans the entire sample, while records of the treasury bill's interest rate only start in 1920. Nevertheless, it is more likely that periods close to the zero lower bound are correlated with periods of high unemployment rates. In fact, this correlation is quite striking as figure 8 shows. Points in orange denote RZ's definition of the zero lower bound dummy. Without controlling for unemployment rates, RZ's estimates for the state-dependence of fiscal policy in times of a binding zero lower bound might just pick

Figure 8: Unemployment and Interest Rates



Note: A scatter plot of unemployment and interest rates. The orange dots depict observations during a ZLB period as defined by RZ.

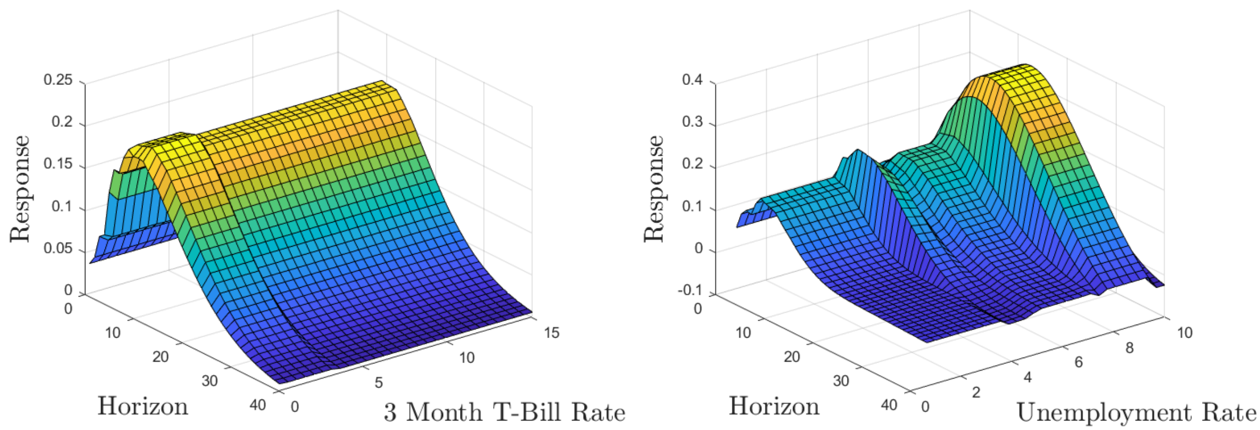
up the effect of high unemployment due to an omitted variable bias.

Another way of visualizing high-dimensional impulse response functions are partial dependency plots. Partial dependency plots are a common tool to make machine learning output easier to interpret. The concept aims to illustrate the marginal effects of individual variables when holding all other variables constant. We make use of this concept and plot partial dependency IRFs in Figure 9a fixing the other variable at its median. The left plot of Figure 9a shows the partial dependency impulse response for variations in the interest rate of the treasury bill. The impulse response appears constant and only slightly higher for interest rates close to the zero lower bound.

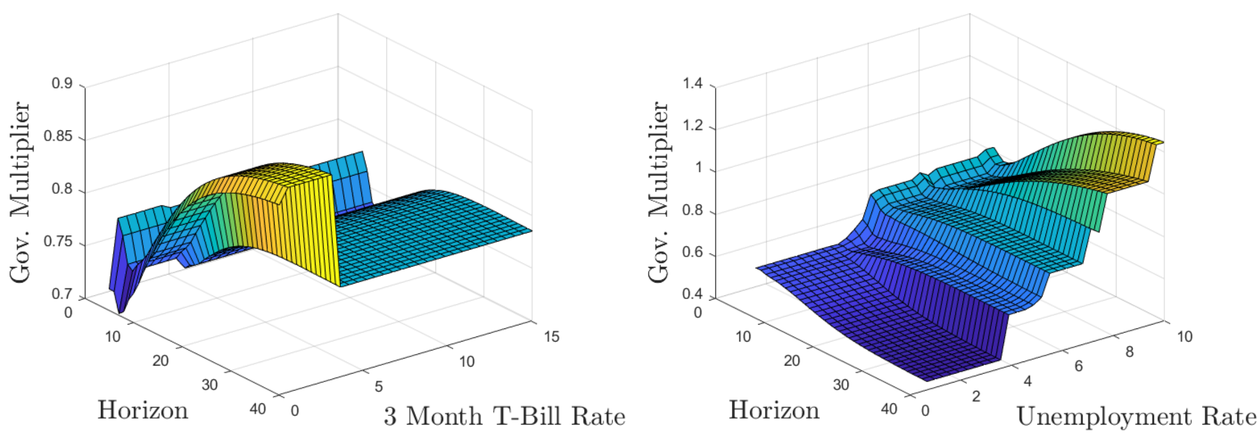
In contrast, the state-dependence of fiscal policy with respect to the unemployment rate is stark. The magnitude of the response of output in case of unemployment rates higher than 8% is more than twice as large as for low unemployment rates. Figure 9b relates the high-dimensional

Figure 9: High-dimensional IRFs

(a) Partial dependency impulse responses

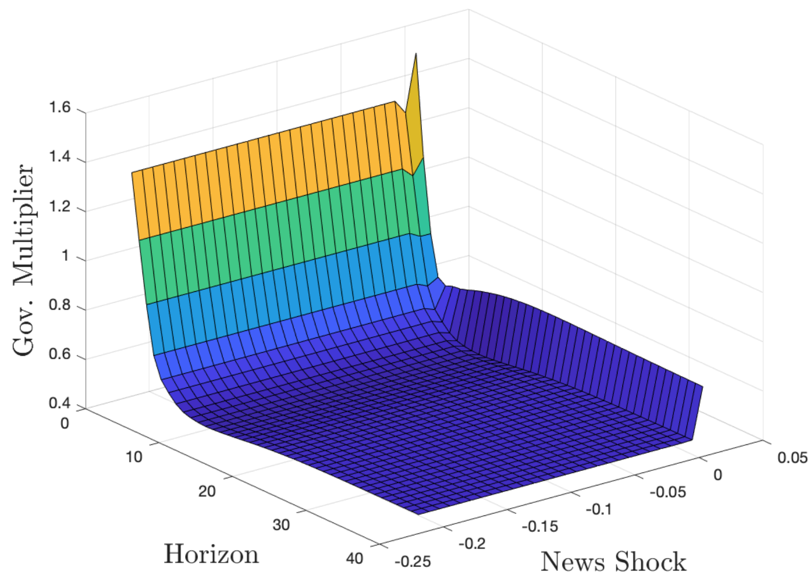


(b) Partial dependency fiscal multiplier



Note: High-dimensional IRFs of output. The top-panel depicts how the moderator variables T-Bill rate and unemployment rate affect the IRF on output. The lower panel analyses the government multiplier.

Figure 10: Fiscal Multiplier for Negative and Positive Gov. Shocks



Note: This figure shows the output multiplier of government spending. To test for asymmetric effects of positive or negative shocks, the News shock series is a moderator variable.

response of output to responses in government spending by calculating high-dimensional fiscal multipliers. We follow Ramey & Zubairy (2018) in computing the multiplier as the cumulated response of output over the cumulated response of government spending[11]. While the multiplier is only slightly higher close to the zero lower bound relative to environments of higher interest rates and below unity, the multiplier is increasing for higher unemployment rates and becomes larger than unity.

4.3 Estimating Asymmetric Effects

As another illustration of the methodology, we assess potential asymmetric policy effects. To allow for the estimation of asymmetric policy effects, the News shock variable (i.e. RZ's instrument for government spending) will now be introduced as a moderator variable. This allows to predict IRFs for different, including negative, values of the shock variable. Note that when

[11] This is the definition Ramey & Zubairy (2018) used to calculate multipliers based on their TVAR estimates.

we include the shock series as a moderator variable, it has to be excluded when estimating the reduced or structural form equation for the news shock series. Otherwise, the variable would appear on both, the left- and right-hand side of the regression equation.

Figure 10 depicts the government multiplier for various shock values. The multiplier is lower for negative as compared to positive values. Furthermore, it appears that the multiplier does not change much over the negative values of the shock series. However, this might be due to the fact that the news series contains few negative observations, making our estimation less precise. Nevertheless, this multiplier appears similar to the ZLB-dependent multipliers from the TVAR in Figure 6.

4.4 Many Moderator Variables

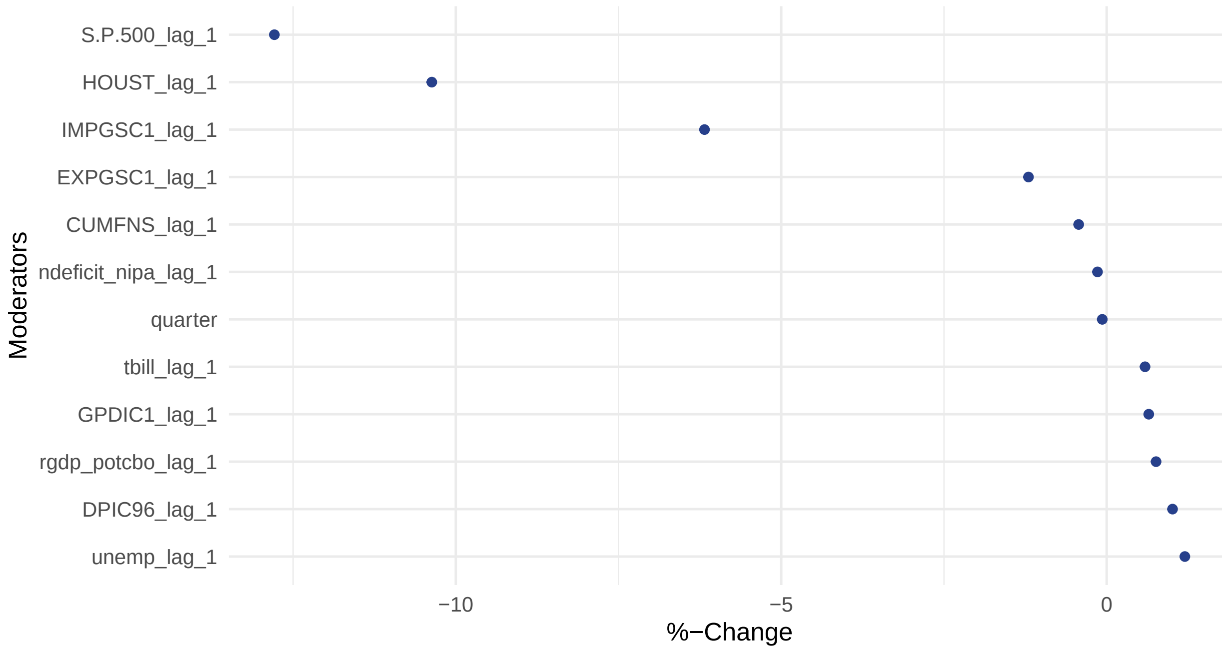
As a final illustration, we include thirteen moderator variables in the previous VC-VAR. With the possibility to include many moderator variables, researchers can be more agnostic with their hypotheses as to which variables might affect the efficacy of economic policies. The variables we include are macroeconomic time series from the FRED data base shown in Appendix B, Table 1.

When including many moderator variables, analysing all possible IRFs becomes infeasible. Instead, we suggest to use the VIM explained in the simulation section to find the most important moderator variables. Figure 11 shows the VIM for the thirteen moderator variables we included in the estimation of the structural equation of output, y[12]. The x-axis depicts the percentage change of the mean squared prediction error when comparing the model with one-variable left out to the full model. For example, estimating the full model as compared to randomizing the observations of *S.P.500_lag_1* decreases the mean squared error by approximately 13%. In other words, including Variable *S.P.500_lag_1* has the biggest impact on improving the mean squared error, which indicates a high importance of this moderator variable.

The unemployment variable appears less important compared to some other moderators.

[12]We get very similar results for the reduced form of output, y (see Appendix C and Figure 18)

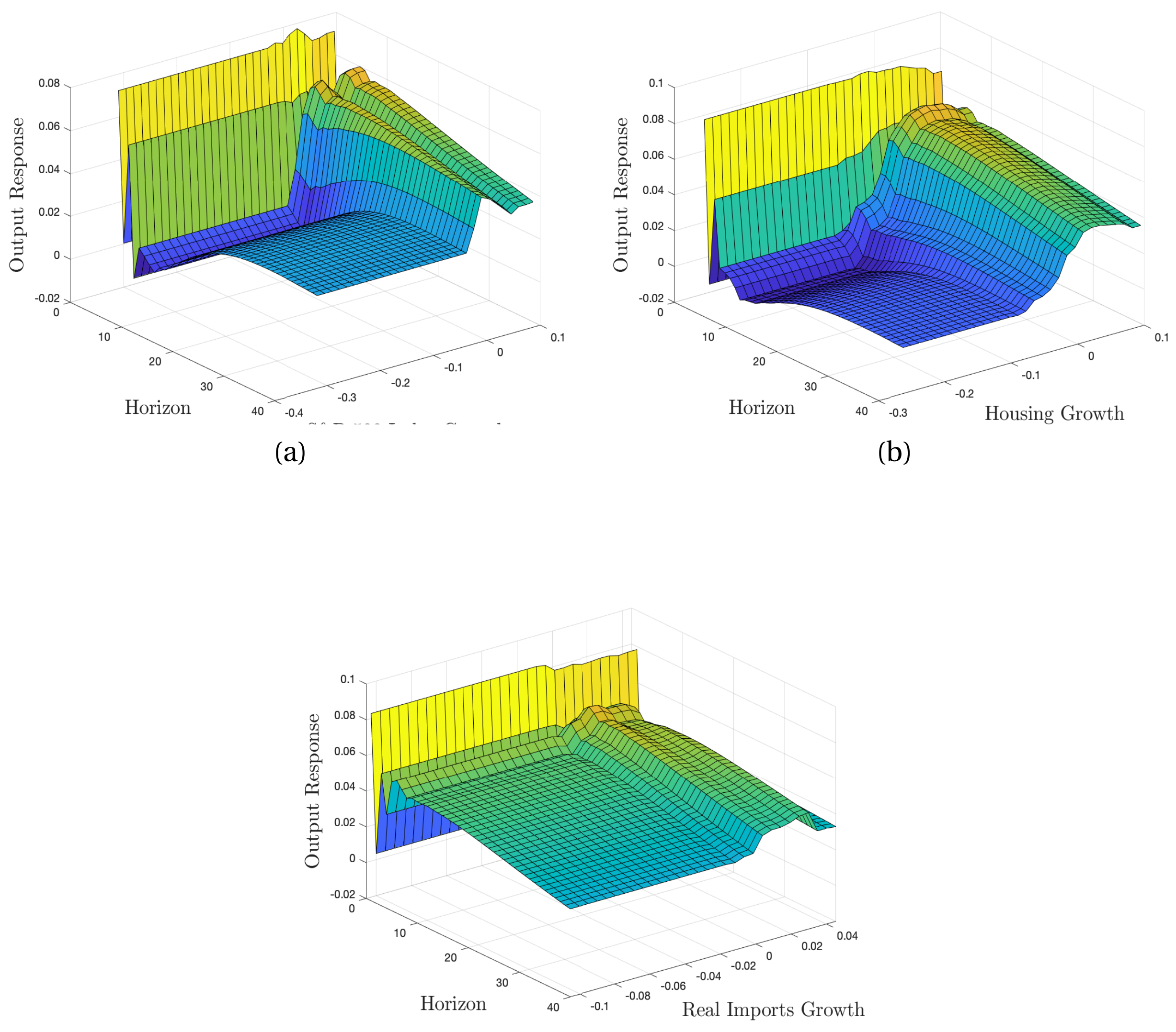
Figure 11: Variable Importance Measure



Note: This figure shows the VIM for the moderator variables used in the estimation of the structural equation of output, y. The greater the percentage change, the more important is the moderator for the estimation.

However, at the moment, the log-likelihood reduction in Algorithm 1 is not standardized to differences in the number of candidate splits (Buergin & Ritschard, 2017). Hence, Algorithm 1 tends to choose moderator variables with many candidate splits more often. The unemployment moderator, *unemp_lag_1*, has 66 and, for example, *S.P.500_lag_1* has 244 unique values out of 244 observations, which might explain some of the differences. Reducing this bias is a focus for future research. Furthermore, it can be seen from Figure 20 in Appendix E that in fact, unemployment does not seem to have a great impact on the output response. However, we also estimate a TVAR on the same time period using the unemployment variable to split the sample between high and low unemployment regimes (see Appendix E, Figure 21). The TVAR shows that when used on its own, the unemployment variable gives rise to differing IRFs and multiplier estimates. Hence, it might be that unemployment is correlated with some of the other moderator variables used in the VC-VAR and combined with the fact that unemployment splits less often, the variable appears less important.

Figure 12: High-Dimensional IRFs Many Moderators



Note: High-dimensional IRFs for output given a 1% news shock and different moderator variables. The estimation was conducted on a shorter sample period, 1960 – 2016.

Figure 12 shows the high-dimensional IRFs for output with the most important moderators according to the VIM in Figure 11. All IRFs show large movements in the first quarters with an impact around 0.08 that quickly drops to then rise again. While this movement appears striking, it is not due to the estimator. The same can be observed when conducting a standard VAR estimation following RZ (2018) on the 1960 – 2016 period (see Appendix D, Figure 19). Similarly, the magnitude of the output response is smaller compared to when estimated on the longer sample period. Again, a similar size of the output response is found when estimating a simple SVAR following RZ (2018) on this shorter time-period (see Appendix D, Figure 19). Hence, the overall dynamics and the magnitude of the output response are in line with standard SVAR estimation results.

Analysing the IRF for the three selected moderators variables, we see that indeed, growth in the S&P 500 Index (Figure 12) as well as in new privately owned housing units (Figure 12) has a strong impact on the output response. For both, S&P 500 and housing, the output response is near zero whenever growth is negative. For positive growth rates, however, the output response peaks at roughly 0.06. Positive growth in the S&P 500 index might indicate a more positive economic outlook from private agents who are in turn more willing to consume and invest the extra income from government spending. Similarly, positive growth in new housing units might indicate a more active economy where government spending will be used for further investments.

Imports in Figure 12 do not appear to give rise to differing output responses. However, the output response appears to be generally on a higher level around 0.04. A small increase in the output response can be observed for positive growth in real imports. Nevertheless, the difference is not as striking as for S&P 500 and housing, which further validates the usage of our VIMs.

5 Conclusion

This paper proposes a semi-parametric framework for estimating high-dimensional SVARs and impulse response functions. While a parametric shell keeps the estimation efficient and more interpretable, tree-based learners allow for non-parametric variation in coefficients depending on potential modifiers. Using tree-based learners has two main advantages: 1) it allows us to detect any potential state-dependence without specifying any functional form or cutoff thresholds, 2) forests can cope with a high-dimensional state space allowing us to control for many potential effect modifiers. Simulation studies confirm that the semi-parametric framework correctly identifies state-dependent variation in coefficients even for macroeconomic samples sizes. When applying the estimator to the historical data set by Ramey & Zubairy (2018) we confirm their findings that fiscal policy is more effective in times of high unemployment. However, we only find slight differences in the efficacy of fiscal policy close to the zero lower bound relative to normal times. In fact, RZ's results seem to be driven by the correlation between low-interest rates and unemployment.

References

Ascari, G. & Haber, T. (2019). *Sticky prices and the transmission mechanism of monetary policy: A minimal test of New Keynesian models.* Technical report.

Auerbach, A. J. & Gorodnichenko, Y. (2012). Fiscal multipliers in recession and expansion. In *Fiscal policy after the financial crisis* (pp. 63–98). University of Chicago Press.

Auerbach, A. J. & Gorodnichenko, Y. (2013). Output spillovers from fiscal policy. *American Economic Review*, 103(3), 141–46.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.

Buergin, R. A. & Ritschard, G. (2017). Coefficient-wise tree-based varying coefficient regression with vcrpart. *Journal of Statistical Software*, 80(6), 1–33.

Canzoneri, M., Collard, F., Dellas, H., & Diba, B. (2016). Fiscal multipliers in recessions. *The Economic Journal*, 126(590), 75–108.

Duarte, V. (2018). Machine learning for continuous-time economics. *Available at SSRN 3012602*.

Eichenbaum, M., Rebelo, S., & Wong, A. (2018). *State dependent effects of monetary policy: The refinancing channel.* Technical report, National Bureau of Economic Research.

Farhi, E. & Werning, I. (2016). Fiscal multipliers: Liquidity traps and currency unions. In *Handbook of macroeconomics*, volume 2 (pp. 2417–2492). Elsevier.

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

Goulet Coulombe, P. (2020). The macroeconomy as a random forest. *Available at SSRN 3633110*.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–779.

Joseph, A. (2019). Shapley regressions: A framework for statistical inference on machine learning models. *arXiv preprint arXiv:1903.04209*.

Li, D., Plagborg-Møller, M., & Wolf, C. K. (2021). Local Projections vs. VARs: Lessons From Thousands of DGPs. (pp.$\tilde{4}$4).

Nakamura, E. & Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3), 59–86.

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics* *and Statistics*, 102(4), 690–704.

Plagborg-Møller, M. & Wolf, C. K. (2019). Local projections and vars estimate the same impulse responses. *Unpublished paper: Department of Economics, Princeton University*, 1.

Ramey, V. A. & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy*, 126(2), 850–901.

Rendahl, P. (2016). Fiscal policy in an unemployment crisis. *The Review of Economic Studies*, 83(3), 1189–1224.

Wang, J. C. & Hastie, T. (2014). Boosted varying-coefficient regression models for product demand prediction. *Journal of Computational and Graphical Statistics*, 23(2), 361–382.

Wolf, C. K. (2020). SVAR (Mis)Identification and the Real Effects of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, 12, 1–32.

Zhou, Y. & Hooker, G. (2019). Tree boosted varying coefficient models. *arXiv preprint arXiv:1904.01058*.

A Simulation studies

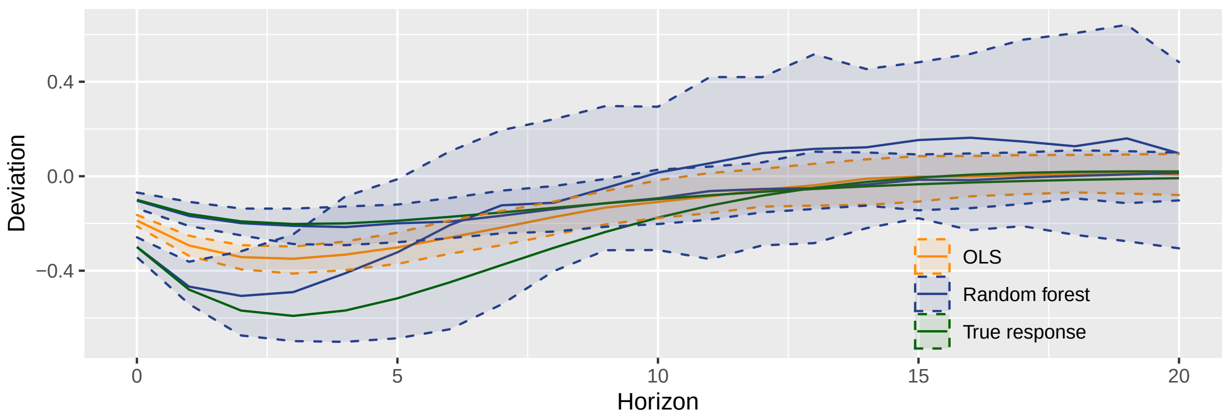


Figure 13: Impulse response estimates for setting 1 using local projections

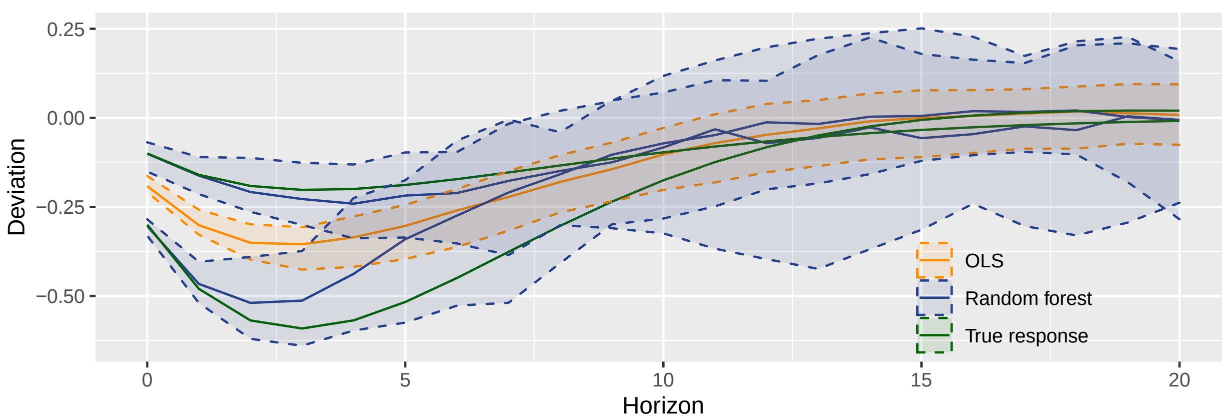


Figure 14: Impulse response estimates for setting 2 using local projections

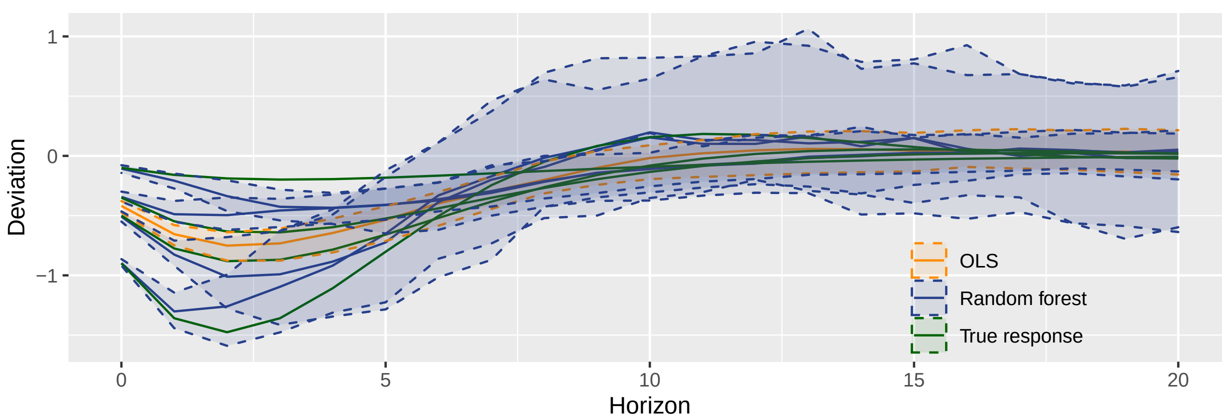


Figure 15: Impulse response estimates for setting 3 using local projections

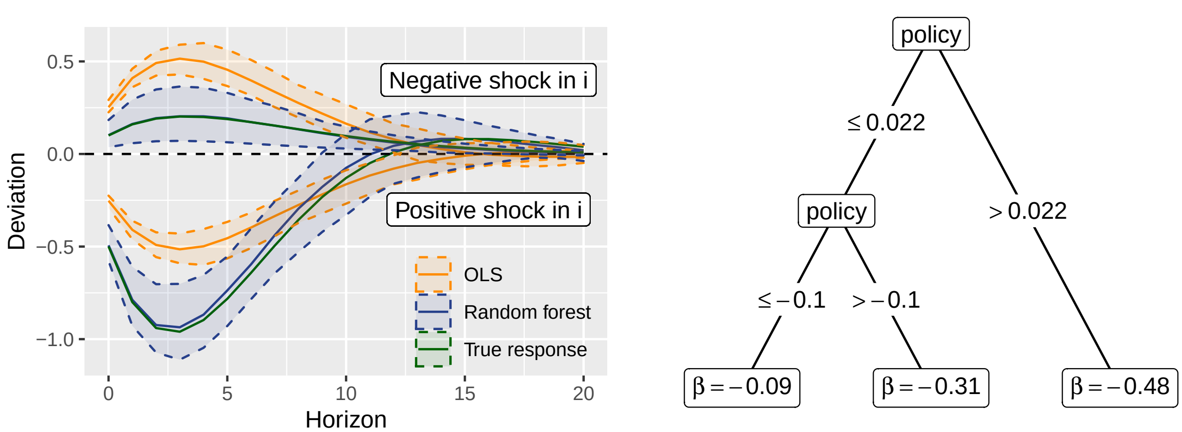
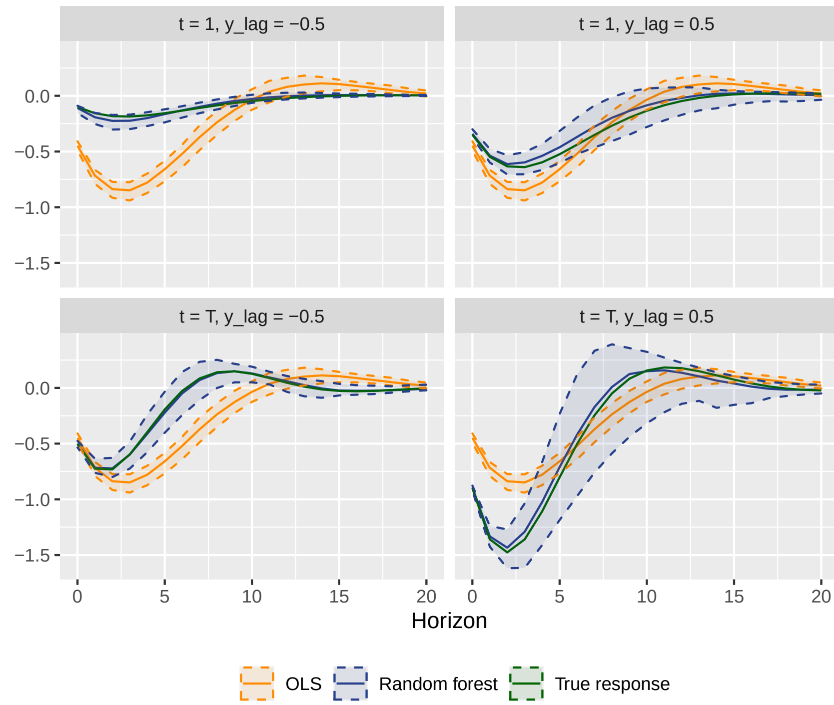
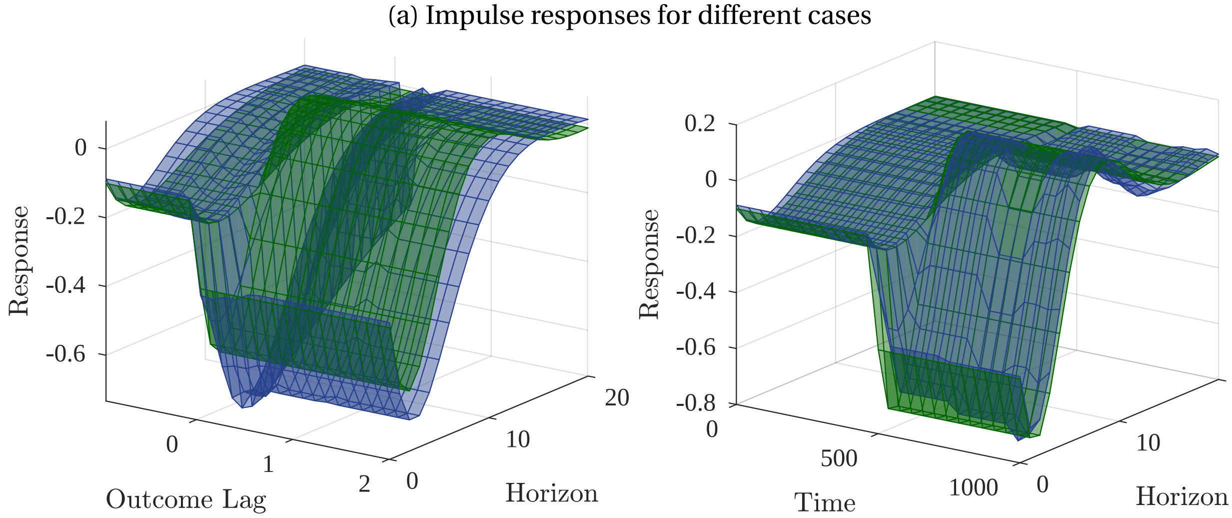


Figure 16: Impulse response estimates in case of an asymmetric policy effect



(a) Impulse responses for different cases



(b) Partial dependency impulse responses

Figure 17: Impulse response functions and their partial dependency for impact and dynamic changes, for T = 1000.

B Moderator Description

This appendix describes the moderator variables used in the empirical section.

Table 1: Description of Moderators

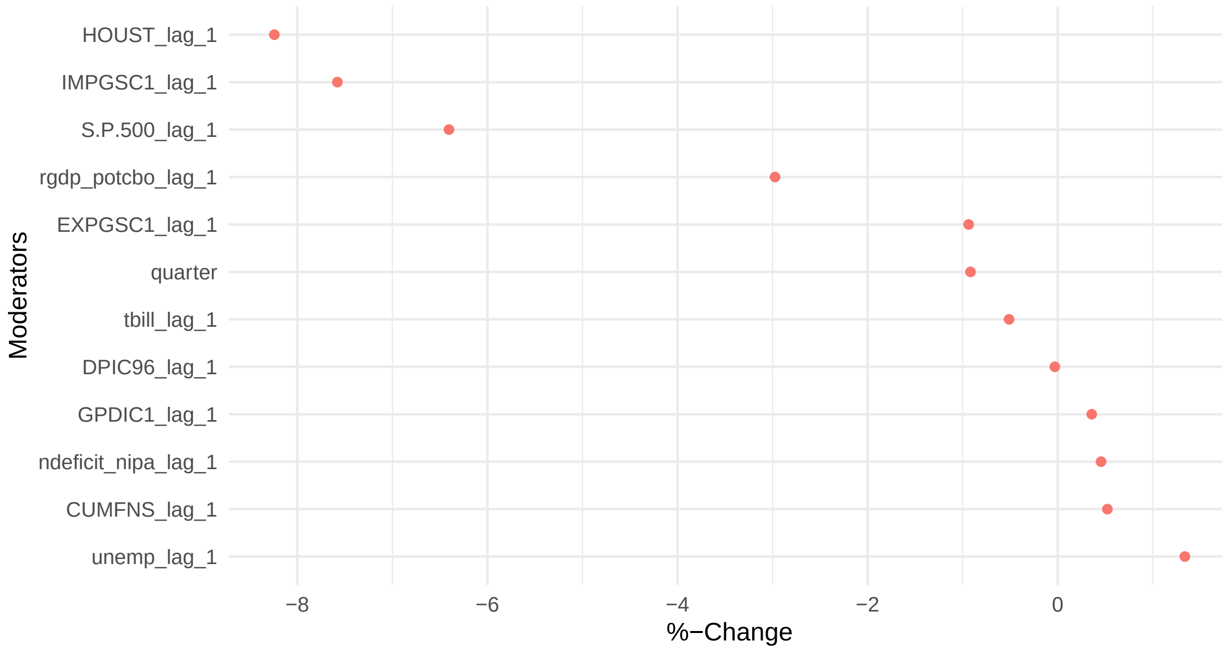
Code	Definition	Source
tbill	3-month treasury bill rate, secondary market.	Ramey and Zubairy (2018).
unemp	Civilian unemployment rate.	Ramey and Zubairy (2018).
wwii	World War II Dummy.	Ramey and Zubairy (2018).
ndeficit_nipa	Nominal deficit, NIPA accrual basis, constructed as total federal expenditures - total federal receipts.	Ramey and Zubairy (2018).
rgdp_potcbo	Real potential GDP, baseod on cubic polynomial fit.	Ramey and Zubairy (2018).
GDPIC1	Real gross private domestic investment. First difference of natural log.	FRED Database (NIPA)
EXPGSC1	Real exports of goods and services. First difference of natural log.	FRED Database (NIPA)
IMPGSC1	Real imports of goods and services. First difference of natural log.	FRED Database (NIPA)
DPIC96	Real disposable personal income. First difference of natural log.	FRED Database (NIPA)
CUMFNS	Capacity utilization: Manufacturing. Percent of capacity.	FRED Database
HOUST	Total new privately owned housing units started. First difference of natural log.	FRED Database
S.P.500	S&P 500 common stock price index: Composite. First difference of natural log.	FRED Database

Note: This table describes the moderators used in the empirical section and their sources.

C VIM Reduced Form Output

This appendix section depicts the VIM for the reduced form estimation of output, y.

Figure 18: IRF from SVAR: 1960-2016



Note: This figure shows the VIM for the moderator variables used in the estimation of the reduced form equation of output, y. The greater the percentage change, the more important is the moderator for the estimation.

D Standard IRF: 1960-2016

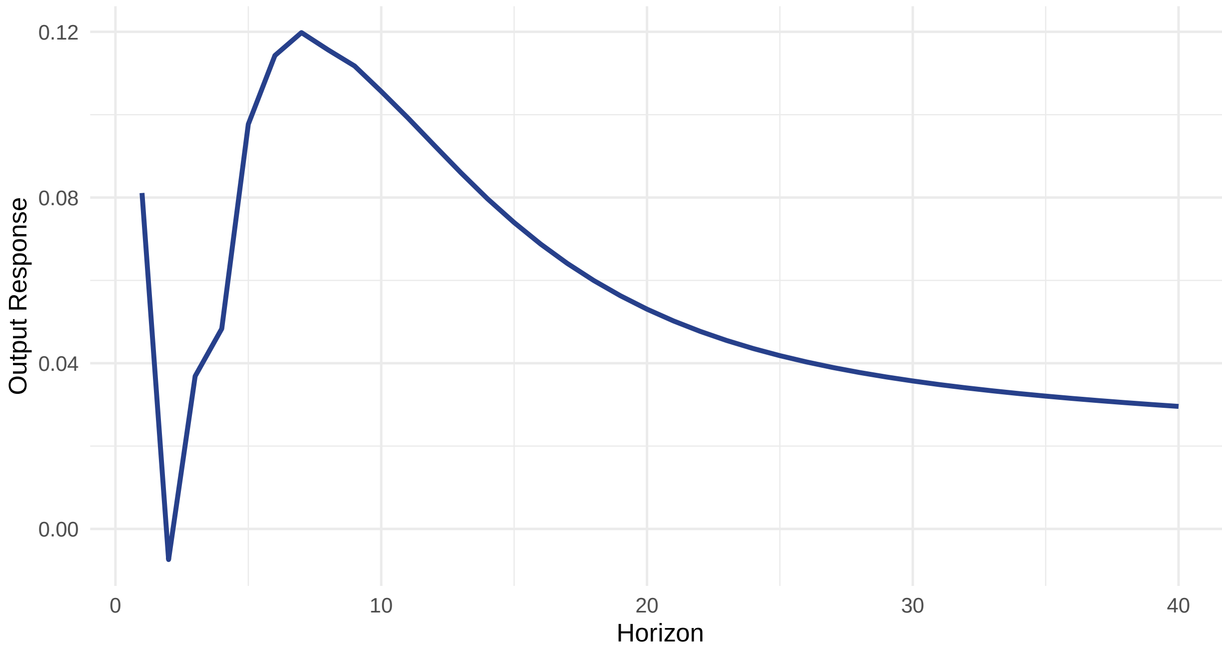
This appendix shows the IRF from a simple VAR estimation following RZ (2018) but for the shorter time period $1960-2016$. From Figure 19 it can be seen that the magnitude of the output response as well as the initial volatility is in-line with what is found for the RF estimation.

E Unemployment as Moderator: 1960-2016

This appendix section shows the IRF for output and unemployment as a moderator while holding all other moderators at their median. Figure 20 shows that compared to the longer sample, different unemployment rates do not seem to affect the output response.

However, when we do a similar exercise using unemployment as the sole moderator variable in a TVAR, we find that IRFs and multiplier estimates differe for high and low unemployment. This might indicate that unemployment is correlated with one or more other moderator variables in the VC-VAR and for this reason, appears to be unimportant.

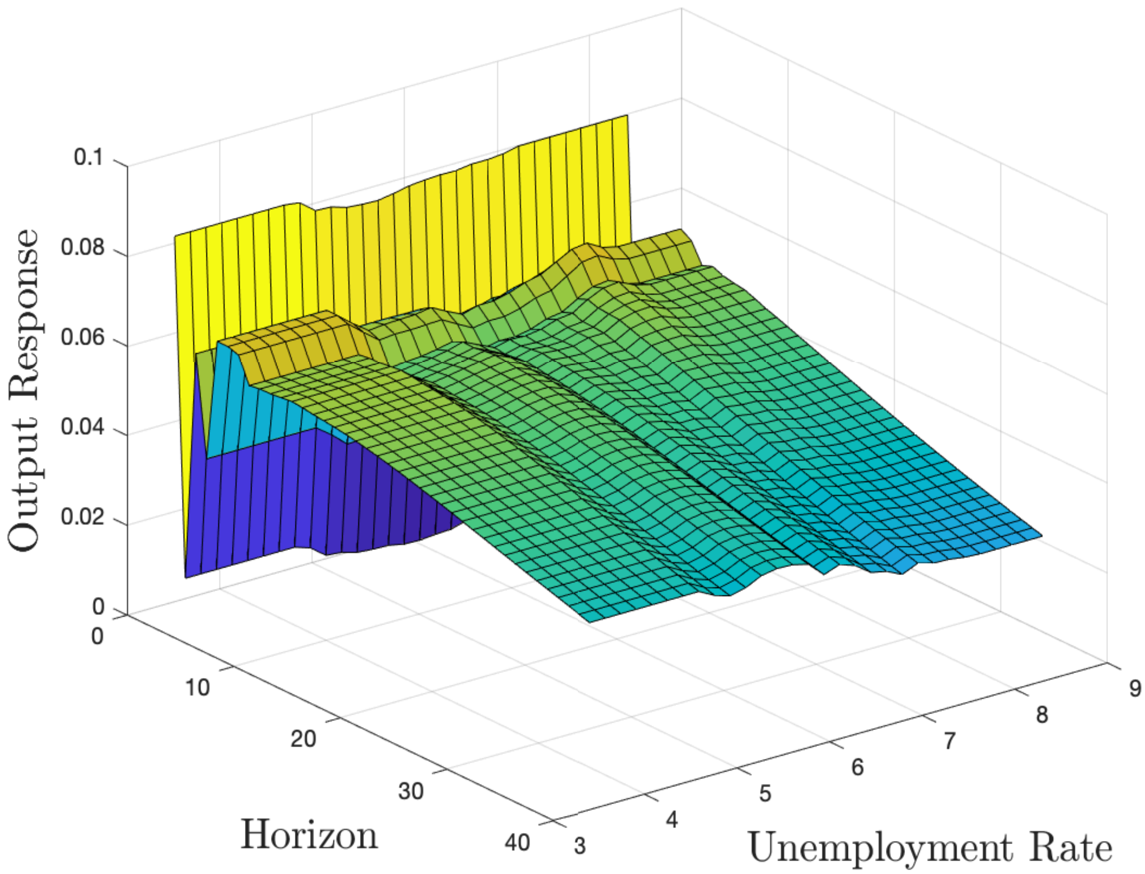
Figure 19: IRF from SVAR: 1960-2016



Note: This figure shows the IRF for output derived from a standard SVAR estimation. The estimation period was 1960 – 2016.

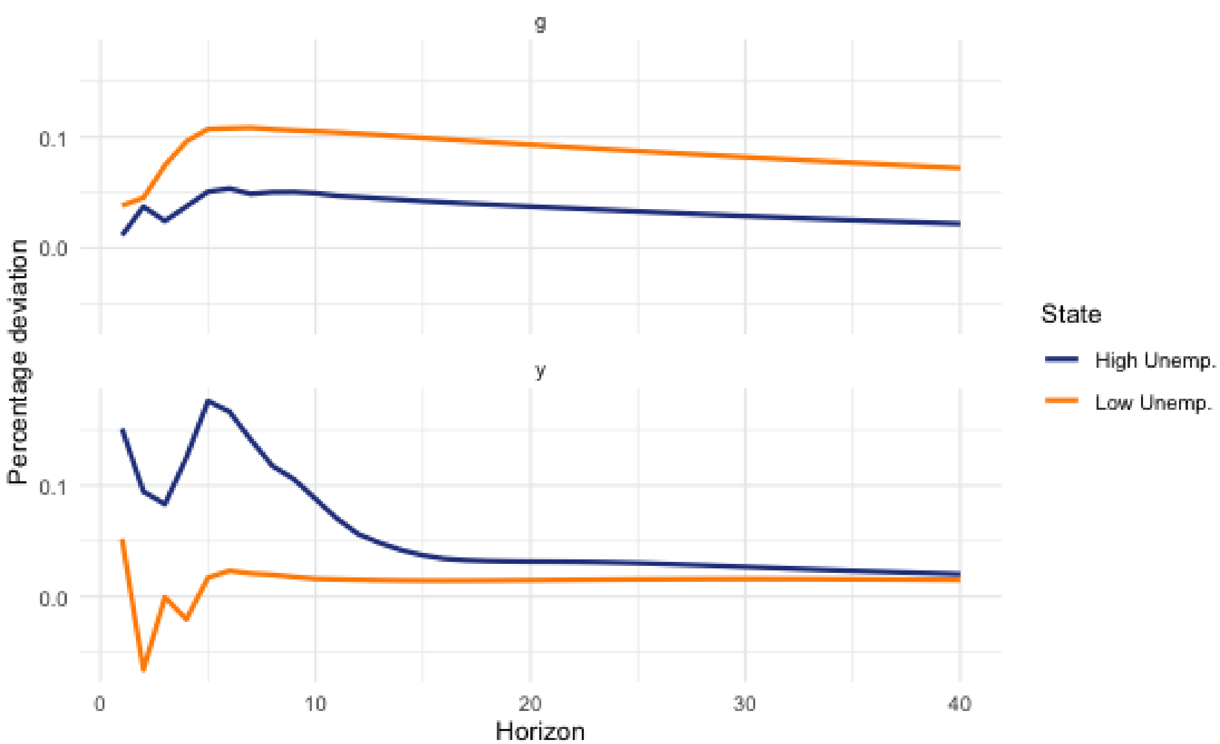
F Partial Dependency Plots

Figure 20: Unemployment as Moderator: 1960-2016

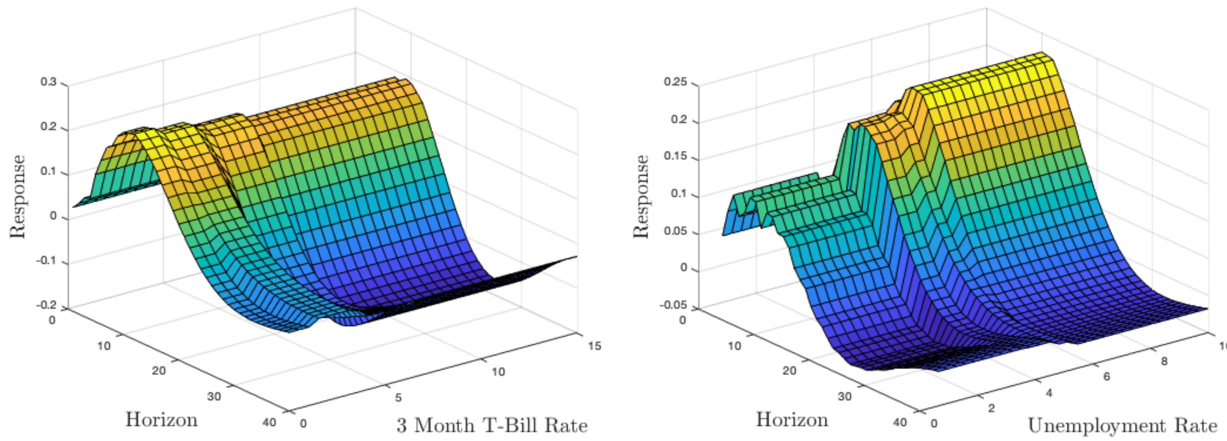


Note: This figure shows the IRF for output and unemployment as the moderator. The estimation period was 1960 – 2016.

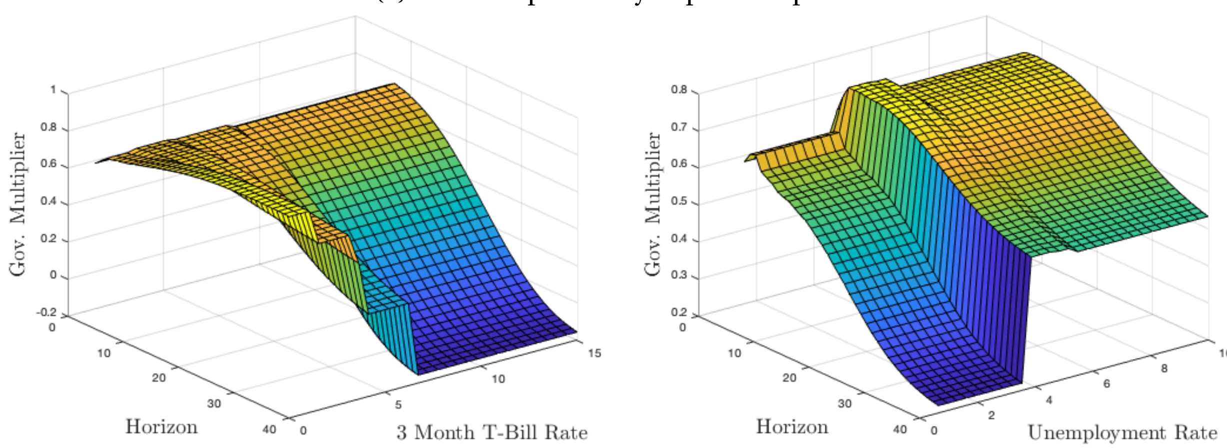
Figure 21: Unemployment as Moderator: 1960-2016



Note: This figure shows the IRF for output with unemployment as the sole moderator in a TVAR estimation. The estimation period was 1960 – 2016.



(a) Partial dependency impulse responses



(b) Partial dependency fiscal multiplier

Figure 22: Estimated partial dependency plots for a varying-coefficient VAR using eight lags.