ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Impulse Response Functions for Self-Exciting Nonlinear Models

Authors	Neville Francis, Michael T. Owyang, and Daniel Soques
Working Paper Number	2023-021B
Revision Date	August 2024
Citable Link	https://doi.org/10.20955/wp.2023.021
Suggested Citation	Francis, N., Owyang, M.T., Soques, D., 2024; Impulse Response Functions for Self-Exciting Nonlinear Models, Federal Reserve Bank of St. Louis Working Paper 2023-021. URL https://doi.org/10.20955/wp.2023.021

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Impulse Response Functions for Self-Exciting Nonlinear Models*

Neville Francis† Michael T. Owyang‡ Daniel Soques§

August 26, 2024

Abstract

We calculate impulse response functions from regime-switching models where the driving variable can respond to the shock. Our focus is on nonlinear vector autoregressions with a variety of specifications for the transition function used throughout the literature. Using Monte Carlo simulations with different misspecifications, we identify the conditions under which impulse response function estimates exhibit significant bias. Furthermore, we extend the concept of model-average impulse responses to this nonlinear context and demonstrate their robustness to model misspecification. Applying these methodologies to the empirical estimation of regime-dependent fiscal multipliers, we find that the multipliers are generally less than one, with small differences observed across varying states of economic slack.

Keywords: generalized impulse response functions, threshold models, regime-switching models, model averaging

JEL Codes: C22, C24, E62

---

*The authors benefited from helpful comments from Graham Elliott, Ana María Herrera, Òscar Jordà, Valerie Ramey, seminar participants at Northeastern University, Indiana University, Kansas University, and participants at the CFE 2021 Conference. The views expressed here are the authors' alone and do not reflect the opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Department of Economics, University of North Carolina at Chapel Hill and NBER

‡Corresponding author. Research Division, Federal Reserve Bank of St. Louis. michael.t.owyang@stls.frb.org

§Department of Economics and Finance, University of North Carolina Wilmington

1 Introduction

The literature on the state-dependent effects of macroeconomic policy highlights how specification choice can affect the economic interpretation of the result. In some regime-switching models, the state of the business cycle affects the underlying macroeconomic dynamics, but shocks also affect economic conditions and, as a result, the state of the business cycle [Auerbach and Gorodnichenko (2012, 2013); Fazzari et al. (2015, 2021); Owyang et al. (2013); Ramey and Zubairy (2018); Laumer and Phillips (2020), among many others]. The feedback in these so-called self-exciting models complicate impulse response function (IRF) computation, as the IRFs depend on the historical data, the past states, and the size and scale of the shock. IRF computation is further complicated by misspecification of the transition process since shocks may not propagate properly into future regime changes.

One approach to address these challenges is estimating generalized impulse responses functions [GIRFs, Koop et al. (1996)], which construct nonlinear IRFs using Monte Carlo methods.[1] Future regimes are determined at each response horizon by simulating possible paths for the variable(s) that determine the state of the economy. An alternative to GIRFs is regime-dependent local projections [RD-LP, Jordà (2005)], which constructs the IRF as the differences between direct multistep forecasts that depend on the state of the economy at the time of the shock [applied examples include Auerbach and Gorodnichenko (2013); Owyang et al. (2013); Ramey and Zubairy (2018, hereafter RZ) among many others]. Local projections have become a popular alternative to standard, iterative IRFs because of their computational simplicity and, in linear models, their robustness to some forms of misspecification [Montiel Olea et al. (2024)]. By projecting onto forward data directly, RD-LP accounts for regime change that actually occurred in the data but does not account for counterfactual switches induced by the policy experiment.

Neither of these methods, however, are a panacea. The GIRFs can be sensitive to the

[1] Another option found in the literature computes IRFs holding the state of the economy fixed throughout the response horizon [e.g., Auerbach and Gorodnichenko (2012)]. Fixing the regimes, however, may overestimate the duration of cyclical downturns in the presence of expansionary policy shocks.

nonlinear VAR's specification, particularly the transition function and the variable(s) driving the regime-switching. For example, using capacity utilization as a transition variable, Fazzari et al. (2015) find government spending multipliers greater than 1 during recessions but smaller than 1 during expansions. In contrast, using the unemployment rate as a transition variable, RZ find multipliers less than 1 in all regimes. The former result incentivizes policymakers to engage in opportunistic spending, while the latter result suggest that spending crowds out the private sector in either business cycle phase. While RD-LP allows the user to estimate the nonlinear VAR without taking a stand on the form of the transition function, it may be subject to other biases. For example, Gonçalves et al. (2023) show that RD-LP is biased in self-exciting models when the shock size is larger than infinitesimal.

In this paper, we investigate the accuracy of IRFs for various self-exciting VARs, focusing on the specification of the transition functions. We identify three popular specifications for the transition functions throughout the literature: (i) the Threshold VAR (TVAR), (ii) the Time-Varying Transition Probability Markov-Switching VAR (MSVAR), and (iii) the Smooth Transition VAR (STVAR).[2]

To address this model uncertainty, we propose using model-average GIRFs (MA-GIRFs) that use weights based on the BIC. The literature on model averaging as a solution to model uncertainty when computing IRFs is sparse. Ho et al. (2023) averages over a variety of structural models using prediction pool in the spirit of Geweke and Amisano (2011), while Li et al. (2021) constructs averages using the Stein combination estimator proposed in Hansen (2016). Our approach extends this literature to self-exciting nonlinear models and uses in-sample fit for weighting, which limits us to maintaining constant model weight across horizons.

Based on simulation exercises, we arrive at a number of conclusions. First, if one knows the true model, one should estimate GIRFs from the true model. Second, in the presence of uncertainty about the transition process, the MA-GIRF that weights all the GIRFs from

[2]For example, Balke (2000) and Ferraresi et al. (2015) estimate a TVAR with credit as the threshold variable and find differences in the responses when credit is tight.

possible models dominates the use of any single model. Third, the use of the MA-GIRF requires using a sufficient number of VAR lags to avoid truncating the model.

We then estimate regime-dependent responses of macroeconomic variables to an increase in government spending from various nonlinear VARs, RD-LP, and the MA-GIRF using U.S. data. Accounting for future state dynamics can be vital to the estimation of fiscal multipliers, as noted by RZ and Caggiano et al. (2015). We use the IRFs to compute fiscal multipliers across states of the business cycle and find relatively small multipliers (less than one) in both slack and nonslack phases across most model specifications. We find scant evidence that the fiscal spending multiplier depends upon the degree of economic slack. Our results suggest that previous studies that find regime-dependent multipliers likely are not robust to model misspecification or alternative shock identification.

The balance of the paper is outlined as follows: Section 2 presents the class of models we study and the methods we will use to compute the impulse responses. Section 3 describes the Monte Carlo experimental design, the metrics used to gauge model-performance, and the simulation results. Section 4 applies the various methods to estimate regime-dependent fiscal multipliers. Section 5 concludes.

2 Methodology

Computing IRFs in a linear setting is straightforward. Nonlinearities, however, can manifest in a variety of ways, each of which may present different problems for the computation of IRFs. To limit the scope of our investigation, we restrict the class of models to self-exciting regime-switching VAR models that vary based on the specification of their state transition processes. A unifying characteristic of the models is that the variable driving the switching process—hereafter denoted z_t—will be an element of the VAR and respond to the shock of interest. To remove one source of uncertainty, we assume that the structural shock, u_t, is observed.

2.1 Self-Exciting Models

We consider three models: (i) the Threshold VAR (TVAR), (ii) the Time-Varying Transition Probability Markov-Switching VAR (MSVAR), and (iii) the Smooth Transition VAR (STVAR). These models are regime-switching, produce regime-dependent impulse responses, and are common in the literature. To fix ideas, a vector of endogenous variables, Y_t, propagates according to:

$$Y_t = (1 - S_t)\, Y_{0t} + S_t Y_{1t} + \varepsilon_t, \tag{1}$$

where the state $S_t \in \{0, 1\}$ in the case of the TVAR and the MSVAR; $S_t = [0, 1]$ in the case of the STVAR; $\varepsilon_t \sim N\,(0, \Sigma_t)$; and Σ_t is the regime-dependent variance-covariance matrix:

$$\Sigma_t = (1 - S_t)\, \Sigma_0 + S_t \Sigma_1. \tag{2}$$

In each state $k \in \{0, 1\}$, the regime-dependent VAR is:

$$Y_{kt} = C_k + B_k(L) Y_{t-1} + D_k\,(L)\, u_t,$$

where C_k is a regime-dependent intercept, $B_k(L)$ and $D_k\,(L)$ are matrix polynomials in the lag operator.

This class of models has several key features. First, the variable governing the state transition process, denoted z_t, is an element of Y_t. Thus, the structural shocks, u_t, that propagate through Y_t produce future changes in regime. Second, the economy takes on two alternative dynamics (or, in the case of the STVAR, a convex combination of two alternative dynamics), dictated by the realization of the underlying state S_t. When $S_t = 0$, the economy has C_0, $B_0\,(L)$, $D_0\,(L)$ dynamics and when $S_t = 1$, the economy has C_1, $B_1\,(L)$, $D_1\,(L)$ dynamics; thus, the model is linear, conditional on S_t being known.

The main difference between the three data generating processes is the specification of the state transition process. In the TVAR, the regime process is determined by the transition variable, z_{t-d} where $d > 0$, relative to a fixed threshold, z^*, as:

$$S_t = \begin{cases} 1 & if \quad z_{t-d} > z^* \\ 0 & if \quad z_{t-d} \leq z^* \end{cases} . \tag{3}$$

The regime dynamics are discrete, meaning that either regime 0 or regime 1 prevails at any time t. The threshold z^* is estimated along with the other model parameters.

The Time-Varying Transition Probability MSVAR [see Diebold et al. (1993) and Filardo (1994) for univariate and Billio et al. (2016) for VARs] has state transition probabilities that depend on a lag of z_t via a logistic transition function, where

$$p_{ji}(z_{t-d}) = \Pr\left[S_t = j | S_{t-1} = i\right] = \frac{\exp\left(\overline{\gamma}_{ji} + \gamma_{ji} z_{t-d}\right)}{\sum_k \exp\left(\overline{\gamma}_{ki} + \gamma_{ki} z_{t-d}\right)} \tag{4}$$

for each of the regimes with $\sum_k p_{ki}\left(z_{t-d}\right) = 1$ for all i, t.

The STVAR [see Granger and Teräsvirta (1993)] has a continuous state process, $S_t\left(z_{t-d}\right) \in [0, 1]$, taking the form of a first-order logistic function:

$$S_t\left(z_{t-d}; \gamma, c\right) = \left[1 + \exp\left(-\gamma\left(z_{t-d} - c\right)\right)\right]^{-1}, \tag{5}$$

where $\gamma \geq 0$ is the speed of transition (As $|\gamma| \to \infty$, the transition becomes sharper and the regime switches resemble a pure threshold model: At $\gamma = 0$, the model collapses to a linear model) and c is a fixed threshold. In equation (5), the regime process is determined by the sign and magnitude of the deviation of z_{t-d} from the estimated threshold c.

2.2 Methods for Computing Impulse Responses

IRFs can be viewed as the difference between two forecasts—one conditional on a structural shock at time t and one conditional on no shock—evaluated over a number of horizons. Leaving aside the problem of identifying the structural shock, IRFs can be based on either iterative multistep forecasts (conventional IRFs computed by propagating a VAR(1)) or direct multistep forecasts (local projections). The literature comparing methods for computing IRFs in linear models is large [Kilian and Kim (2011), Plagborg-Møller and Wolf (2021), Li et al. (2021)]; relatively fewer papers consider self-exciting models [e.g., Gonçalves et al. (2023)]. In this section, we review how GIRFs and RD-LP estimate IRFs in this class of models.[3]

2.2.1 Generalized Impulse Response Functions

One method that accounts for future regime changes is GIRFs [see Koop et al. (1996)].[4] GIRFs use Monte Carlo (MC) methods to draw from (i) the simulated historical paths of the data and (ii) future reduced-form shocks to obtain the difference between one path where a structural shock size, $u_t = \delta$, occurs at time t and a path where no structural shock occurs ($u_t = 0$).[5] To account for the regime switching, we compute the regime-dependent GIRF as the average over the set of histories where $S_{t-1} = i$.

Define $\{\mathbf{Y}_\tau^i\}$ as the set of all histories that correspond to $S_\tau = i$, with $i = 0, 1$. Let $R_1 = \sum_{t=1}^{T} S_t$ represent the number of occurrences of $S_t = 1$ in our dataset and let $R_0 = T - \sum_t S_t$ represent the number of occurrences of $S_t = 0$. Conditional on a history, $\mathbf{Y}_{t-1}^i$, we can simulate two paths for $Y_t, ..., Y_{t+h}$—one based on a structural shock $u_t = \delta$ occurring at time t and one with no structural shock, $u_t = 0$. Both paths are subject to the same path of future reduced-form shocks, $\left\{\varepsilon_{t+l}^{[q]}\right\}_{l=0}^{h}$. We then average over Q MC draws of future shocks

[3] Gonçalves et al. (2023) propose an alternative nonparametric method to estimate the conditional average IRF. They do not, however, estimate IRFs using this alternative or compare it to existing methods. We leave this for future research.

[4] If the transition probabilities are constant, Krolzig (2006) suggests computing a weighted average of the regime-dependent responses, where the weights are determined by the transition probabilities. These models are also called "exogenous" switching models and are explored in detail in Gonçalves et al. (2023). Because our interest is in exploring the effect of the interaction between variables in the VAR and the state process, we forgo analyzing exogenous switching models here. One can, however, think of the exogenous models as a special case of endogenous switching models.

[5] In the linear model, δ is a scaling factor that leaves the shape of the IRF unchanged. In self-exciting models, δ affects the future state process. Size asymmetry is explored in Gonçalves et al. (2023), who show, among other things, that RD-LP is biased for non-Infinitesimal δ. We note, though, that one can easily augment the methods used here with those in Gonçalves et al. (2023).

to obtain the expectation of the two paths, differentiated by the presence of the period$-t$ shock:

$$E_t\left[Y_{t+h}|\Theta,\mathbf{Y}_{t-1}^i,u_t\right]=\frac{1}{Q}\sum_q^Q\left[Y_{t+h}|\Theta,\mathbf{Y}_{t-1}^i,u_t,\left\{\varepsilon_{t+l}^{[q]}\right\}_{l=0}^h\right],$$

where Θ represents the set of model parameters. The value of the GIRF at horizon h after a structural shock of size δ occurs at time t with previous regime $S_{t-1}=i$ is obtained by averaging across all $S_{t-1}=i$ histories:

$$\Phi_{GIRF}^i(h,\delta,\Theta)=\frac{1}{R_i}\sum_{\mathbf{Y}_{t-1}^i\in\left\{\mathbf{Y}_{t-1}^i\right\}}\left\{E_t\left[Y_{t+h}|\Theta,\mathbf{Y}_{t-1}^i,u_t=\delta\right]-E_t\left[Y_{t+h}|\Theta,\mathbf{Y}_{t-1}^i,u_t=0\right]\right\}. \tag{6}$$

Allowing the endogenous variables to start in different states and accounting for the future paths of the driving variable captures future regime switches over the response periods.

2.2.2 Regime-Dependent Local Projections

Unlike GIRFs, RD-LPs do not require parametric assumptions about the transition function, potentially making it more robust to misspecification in nonlinear models [Jordà (2005); Plagborg-Møller and Wolf (2021); Montiel Olea and Plagborg-Møller (2021)].[6] However, Gonçalves et al. (2023) show that RD-LP is biased for self-exciting models when the shock is not infinitesimal [i.e., of the order of shocks used in RZ]. The bias is small if the economy does not change state, suggesting it results from RD-LP's failure to fully capture the shock's effect on the future state. Nevertheless, it may be useful to compare the relative bias of RD-LP with that of GIRFs when the transition function is misspecified.

RZ compute RD-LPs, conditioning on the regime just before the shock incidence, by regressing separately for each horizon h the responding variable on the value of the period$-t$

[6] Kilian and Kim (2011), Plagborg-Møller and Wolf (2021), and Li et al. (2021) among others compare IRFs from VARs and from local projections for linear models.

observed structural shock and a vector of controls:

$$Y_{t+h} = (1 - S_{t-1}) \left(\alpha_{0h} + \beta_{0h} u_t + \gamma_{0h} X_t\right) + S_{t-1} \left(\alpha_{1h} + \beta_{1h} u_t + \gamma_{1h} X_t\right) + v_{t+h},$$

where S_{t-1} is the prevailing regime at the time period before the shock and X_t is the vector of controls.[7] In the simulation exercise, we compute RD-LP using lags of Y_t and lags of the shocks as controls as:

$$\Phi^i_{LP} (h, \delta) = \beta_{ih} \delta,$$

where the horizon-specific coefficient, β_{ih}, depends on the state at the time of the shock. The idea is that the use of Y_{t+h} as the regressand accounts for regime changes that actually occurred in the data between periods t and $t + h$.

2.2.3 Model-Average Generalized Impulse Responses

A method of addressing model uncertainty is the use of model averaging to construct a weighted IRF [see Ho et al. (2023) and Li et al. (2021) for applications with linear models and Steel (2020) for a recent overview of model averaging]. Let $m = 1, .., M$ represent the different models under consideration. The MA-GIRF is:

$$\Phi_{BMA} (h, \delta) = \sum_{m=1}^{M} w_m \Phi_m (h, \delta)$$

where the model weights, w_m, are approximate posterior model probabilities as in Raftery (1995):[8]

$$w_m = \frac{\exp\left(-BIC_m/2\right)}{\sum_{j=1}^{M} \exp\left(-BIC_j/2\right)}. \tag{7}$$

[7] We could estimate RD-LP using the true states, but this gives RD-LP an unfair advantage. Instead, we set the state using a threshold variable, $S_t = I_{[z_t > \bar{z}]}$ relative to its historical average, $\bar{z}$, where $I_{[.]}$ is an indicator function. We consider only discrete S for the starting conditions: Extension to non-discrete S in the spirit of Ruisi (2019), Lusompa (2021), and Inoue et al. (2022) is straightforward.

[8] Hansen (2007) refers to this form of model averaging as Smoothed BIC. Using the BICs to weight the IRFs can lead to better out-of-sample characteristics [Swanson and Zeng (2001)]

Several features of the MA approach proposed here are worth highlighting. Our weights do not vary across the response horizon [as they do in Ho et al. (2023) and Li et al. (2021)] and do not vary across regimes. They are constructed using an in-sample measure of fit, and, thus, preclude use of the RD-LP in the model set. While the weights are applied to the IRFs directly and not to the coefficients, they could be applied directly to any object of interest (e.g., the multiplier in our application). Finally, our weights can also accommodate prior information either on the models or on the impulse responses themselves.

3 Simulation Study

We compare methods of generating IRFs in Monte Carlo exercises where the "true" IRFs are known. Given a data-generating model, at each Monte Carlo iteration, we simulate $\mathbf{Y}_T = \{Y_1, ..., Y_T\}$, $\mathbf{S}_T = \{S_1, ..., S_T\}$, and a time series of exogenous structural shocks. IRFs to the known structural shocks are then computed using the different methods discussed above. The results below are based on 5,000 total simulations for each true DGP – 200 simulated data paths from 25 parameterizations drawn from the posterior distributions of each model.

For each draw of the data, GIRFs are obtained from the parametric models for various forms of misspecification [e.g., lag lengths, conditioning variables, and the driving variable]. At each draw of the Gibbs sampler, $Q = 1,000$ artificial histories are used to compute the GIRFs. For the MA-GIRFs, we set the model weights as in eq. (7) with a flat model prior.

3.1 Defining the True Impulse Responses

In a self-exciting model, solving for the true IRF analytically is intractable [see Gonçalves et al. (2023)]. The true regime-dependent IRFs are obtained using simulation methods, similar to estimating GIRFs as in Koop et al. (1996). The true IRF, $\Phi^i_{GIRF}\left(h, \delta, \widetilde{\Theta}\right)$, is obtained as a GIRF from the correct model specification and true parameters, $\widetilde{\Theta}$. The simulated histories, $\{\mathbf{Y}_t^i\}$, are drawn from a longer sample $Y_{T^L} = \{Y_t\}_{t=1}^{T^L}$ to more accurately approximate the true IRF, while only the last $T < T^L$ observations are used in estimation.

The true parameters of the DGP are calibrated to a fiscal VAR that includes government spending, output, and the Ramey military news shock for the period 1949 to 2015 [see RZ]. We set $T = 250$ and $T^L = 1,000$ in our simulations. We set the size of the shock $\delta = 1$, equal to one percent of GDP based on our calibration setup.

3.2 Performance Metric

We evaluate the estimated IRFs using the Continuous Ranked Probability Score (CRPS, Matheson and Winkler (1976)), either taken as a function of h or averaged over the response horizon, $h = 1, ..., H$.[9] The CRPS compares the fit of the entire posterior distribution with the point-valued true response at each horizon h, using the empirical CDF-based approximation from Krüger et al. (2016) from the MCMC output of each model:

$$CRPS(F, h) = \int_{-\infty}^{\infty} \left[F(\chi) - \mathbb{1}\left(\chi > \Phi_{GIRF}^i\left(h, \delta, \widetilde{\Theta}\right)\right)\right]^2 d\chi,$$

where F is the forecast CDF implied by the posterior distribution of $\Phi_{GIRF}^i(h, \delta, \Theta)$. The indicator function, $\mathbb{1}(\chi > \Phi_{GIRF}^i(h, \delta, \widetilde{\Theta}))$, captures the true distribution, equaling 0 for all χ values up until the true value $\Phi_{GIRF}^i\left(h, \delta, \widetilde{\Theta}\right)$, then 1 for greater values. Therefore, if our posterior were precisely estimated, $F(\chi)$ would follow exactly this pattern. A lower CRPS indicates a better fit as the probabilistic forecast of the CDF is closer to the true response. In comparing one model's overall fit to another, we will consider $CRPS(F)$ which is $CRPS(F, h)$ integrated over impulse response horizons horizons $h = 1, ..., H$.

[9] We also compared the estimated IRFs using the Deviation in State Asymmetry (DSA), which measures the difference in the responses across the initial regime at each horizon. Results for DSA are comparable to those with CRPS and are available in the online appendix.

3.3 Simulation Results

Table 1 shows the CRPSs for the IRFs obtained using various methods, where each panel depicts a different potential source of misspecification. The value reported is the average CRPS over all horizons, $h = 1, ..., 20$. Within a panel, the rows show the true model and the columns show the method used to compute the impulse response. The first three columns provide results for GIRFs using the indicated model; the final two columns show results for RD-LP and the MA-GIRFs weighted across the three GIRFs. In the top panel, the form of the transition process is the only source of misspecification; thus, the diagonal of the top panel represents the use of the true (correct) transition process. Subsequent panels also include misspecification in (i) the lag used for the transition function, (ii) the transition variable, and (iii) the VAR lag.[10] The bottom panel shows results when the true model is a linear VAR but the estimated impulse response are computed assuming state asymmetry.

A broad evaluation of the results suggests several key takeaways. First, except for when the STVAR is the true model, the GIRF from a model estimated with the true transition function generally exhibits less bias than alternative methods. This result is obtained even when misspecifying the transition function lag or variable, so long as the functional form is correct. MA-GIRF consistently performs second best to the true model, suggesting that the effect of model uncertainty can be mitigated by model averaging. RD-LP exhibits more bias than other methods, possibly because it does not directly model the state transition process. Structural shocks that affect the driving variable put the state process on a counterfactual path; RD-LP uses (an average of) empirical paths that can differ from this counterfactual path. Thus, if one is relatively certain about the functional form of the state transition process, one is better off computing GIRFs from estimates of that model.

Second, Figure 1 shows the median CRPS for each model at each h to demonstrate how the relative performance of each model varies over the 20-period horizon. For each model,

[10] For (i), true model uses z_{t-1} as the transition variable but the estimated model uses z_{t-4}. For (ii), the transition variable in the estimated model is uncorrelated with the true z_t. For (iii), the true model is a VAR(4) but the estimated model is a VAR(1).

the bias is smaller at short horizons relative to long horizons, with RD-LP's performance deteriorating faster than the GIRFs as h increases. The latter result likely obtains because RD-LP's approximation of the underlying state worsens as h increases. Thus, estimates of shorter-run effects will be less biased than longer-term estimates. Gonçalves et al. (2023) similarly find in simulations that the asymptotic bias is larger when computing four-year responses compared to two-year estimates.[11]

Third, we note that GIRFs from the estimated STVAR are typically the most biased of the three models, even when the STVAR is the true model. This result suggest that the STVAR should only be used if one has strong prior belief that it is the true model.[12] Conversely, the TVAR—the most sparsely parameterized model of the three—generally performs well, even when it is not the true transition process. Further, the MA-GIRF can perform better than even the true model. Taken together, these results highlight that IRF computation may be thought of as a forecasting problem.

Fourth, truncating the lags in the VAR leads to relatively more bias compared to overfitting with a high lag order. In all cases of VAR lag misspecification, the true model performs the best with MA-GIRF as a close second. However, we note that RD-LP can exhibit less bias than GIRFs when the lag order of the VAR is truncated. The result holds even for RD-LP relative to MA-GIRFs, suggesting that the VAR dynamics themselves—not the form of the transition process—are of first-order interest. Including more lags than the true DGP, however, mitigates the problem for the GIRFs. Thus, we conclude that MA-GIRFs dominate RD-LP and VARs with a misspecified transition function if the dynamics of the true model can be approximated from the model choice set.[13] Thus, prior model uncertainty about the transition process and/or the VAR lag order can be resolved by estimating the MA-GIRFs

[11]The appendix presents CRPS estimates by horizon for the other misspecification cases.

[12]This result might seem puzzling at first as the STVAR nests the TVAR for $\gamma \to \infty$. It may be that estimation of γ introduces enough uncertainty in the transition process that STVAR exhibits more bias. However, if one places substantial prior weight on the true model being linear, the STVAR remains a viable option as it nests the linear model for $\gamma \to 0$.

[13]Note that the MA does not always place the largest weight on the true transition process.

and including a VAR with a high lag order.[14]

Finally, regime-dependent models are often used to determine if an impulse response is larger in one regime compared to another. One common example is whether the fiscal multiplier depends upon the degree of slack in an economy. In results presented in the appendix, we show that a model's ability to capture state asymmetry is in line with its overall performance in terms of bias. The key exception is that the MA-GIRF captures state asymmetry better in all misspecification cases, even when the VAR's lag order is truncated.

4 Application: Fiscal Multiplier

The self-exciting models considered above have a number of macroeconomic applications. Previous studies have used these models to estimate the the regime-dependent effects of uncertainty, monetary policy, and fiscal policy shocks, among others. For the government spending multiplier, in particular, results have varied in important ways. Some studies find that the multiplier is larger in periods of economic slack than in nonslack periods when crowding out is more costly. Other studies find that the multiplier does not vary across states of the business cycle. Figure 2 reviews the regime-dependent multiplier estimates from a small fraction of the studies on fiscal multipliers.[15]

The multiplier is typically computed from the regime-dependent IRFs; thus, differences in assumptions about the underlying VAR model, the transition function, or the driving variable can produce variation in the multiplier. Results from our simulation study suggest that misspecification in either the VAR dynamics or the transition process can create substantial bias in the IRFs in nonlinear models. We consider whether these types of biases might explain the variation in the regime-dependent fiscal multipliers found in the literature. One proposed solution has been to compute the multipliers using RD-LP, which has been claimed

[14] If one suspects the true model dynamics can only be captured by a VAR with a high lag order making the problem computationally difficult, our result suggests that RD-LP would be a suitable alternative.

[15] See Gechert and Rannenberg (2018) for a recent survey on regime-dependent fiscal multipliers. See Ramey (2019) and Castelnuovo and Lim (2019) for recent surveys on the fiscal multiplier in general.

to be more robust to misspecification. We also estimate the MA-GIRFs with a flat model prior, which we showed in the previous section are also robust to model uncertainty.

4.1 Data

The baseline sample, based on RZ, includes real GDP and real government spending and runs from 1890Q1-2015Q4. We also consider a truncated subsample from 1969Q1 to 2015Q4. Output and government spending are scaled by potential output, which is computed as a sixth-order polynomial trend in output [Gordon and Krenn (2010)]. All data are seasonally adjusted.

We consider two commonly-used methods to identify exogenous fiscal shocks: (i) the narrative shock to military spending identified by Ramey (2011) and (ii) the timing assumption from Blanchard and Perotti (2002). We calculate the impulse response of output and government spending to a military news shock or unexpected spending shock equal to 1 percent of GDP. The multiplier is calculated as the cumulative response of output divided by the cumulative response of government spending over a five-year period.

4.2 Results

The top and bottom panels of Table 2 show the cumulative five-year regime-dependent fiscal multiplier estimates for the various models when using the full and short samples, respectively. The first two columns show the multipliers using the Ramey news shock series for the slack and nonslack regimes, respectively. The next two columns show the corresponding results using Blanchard-Perotti (BP) shock identification. Each GIRF is estimated with four lags of Y_t and the output gap as the transition variable. RD-LP is estimated at each horizon, conditional on four lags of Y_t, and requires a definition of the slack and nonslack states across time. We define the states based on whether the output gap is above (nonslack) or below (slack) its historical average.[16]

[16]The results are robust to other choices for the output gap threshold, such as 0.

The results in top panel of Table 2 largely match those found by RZ, Barnichon et al. (2021), and Alloza (2022): (i) multipliers tend to be lower when using the BP identification compared to the Ramey news shocks; (ii) point estimates for the regime-dependent multipliers are both below unity; and (iii) significant differences in multipliers across states can be found using the BP identification. Statistically relevant differences across states are only found when using the BP identification with RD-LP.

The bottom panel of Table 2 shows the results for the shorter sample, which differs in a number of important dimensions from the full sample results. First, the error bands are substantially larger than in the longer sample. This result is consistent with the results of Ramey (2011) and RZ and likely reflects both increased parameter uncertainty and weaker identification of the Ramey news shocks when excluding the Korean War. Second, in the shorter sample, the estimated multipliers occasionally have point values larger than 1. Third, when using RD-LP, the cumulative multipliers implied by the Ramey shocks are now negative, implying a government spending shock contracts the economy. This result matches the negative post-WW2 multipliers of Ramey (2011) and contrasts with the relatively high multipliers from Fisher and Peters (2010).[17] Fourth, as in the full sample, using the BP identification with RD-LP produces multipliers that are statistically higher in slack compared to nonslack states.

The bottom row of each panel shows the multipliers computed using a weighted average across VAR models and identification schemes. These combined MA-GIRF multipliers are slightly larger in the small sample versus the full sample. However, both regime-dependent multipliers are below unity with no distinguishable difference across slack and nonslack states in both samples. Similar to the MA multipliers computed for each identification scheme, the combined MA multipliers have wider error bands when using the shorter sample.

[17] One potential explanation put forth by Ramey (2011) is that the military news shocks are temporary, while other identifications [e.g., Fisher and Peters (2010) shocks to defense stocks' excess returns] are more permanent. Also, a number of papers note that the Ramey news shock series is a poor instrument for government spending shocks after the Korean War, especially compared to shocks identified using Blanchard-Perotti [see Ramey (2016), Kang and Kim (2022), and Jørgensen and Ravn (2022), among others].

Figure 3 shows the regime-dependent MA-GIRFs for output and government spending using the Ramey news shock the full sample. The top row shows the responses of government spending and output in slack states while the bottom row shows the corresponding responses in nonslack states. Despite variation across the nonlinear VAR estimates, the regime-dependent MA multiplier estimates are relatively close to the RD-LP estimates when using the Ramey news shocks. BP identification finds a similar multiplier during nonslack states when using either the MA-GIRFs or RD-LP. However, the MA finds a substantially lower multiplier of 0.46 for slack states compared to the RD-LP estimate of 0.79.

4.3 Robustness

We explore two variations of the baseline specification. First, our simulation results suggest that truncating the VAR lags can induce significant bias in the impulse responses and, subsequently, in the multiplier. Second, much of the heterogeneity in the fiscal multipliers has been attributed to the use of different transition variables. For example, Auerbach and Gorodnichenko (2012) use a seven-quarter moving average of the output gap; RZ use the unemployment rate, NBER recession dates, and the output gap relative to fixed thresholds; Fazzari et al. (2015) uses an adjusted measure of capacity utilization; and Fazzari et al. (2021) uses the output gap. We check the sensitivity of the multiplier estimates to each of these model specifications. Finally, we estimate a "comprehensive multiplier" based on the MA-GIRF over all of the specifications considered.

4.3.1 VAR Lags

The first six columns of Table 3 and Table 4 examine the sensitivity of the estimated multipliers when varying the number of VAR lags. Table 3 considers the full sample whereas Table 4 uses the short sample discussed above. Because of the aforementioned identification concerns with the military news variable when excluding the Korean War, we present only the BP identification. Results for the Ramey news variable are available in the Appendix.

For the long sample, there is variation in the estimated values of the multiplier, but they generally have magnitudes less than one for all lag orders. These results are consistent regardless of the identification of the fiscal shock. In the short sample estimated with more controls, we find more instances of multipliers greater than one for models with VARs with higher lag order.

While there is some variation in the value of the multiplier for different VAR lag orders, we may be able to resolve this uncertainty by examining the MA-GIRFs computed across the various models. This resolution stems from the MA-GIRFs weighting the GIRFs from each given model by their relative fit. The bottom rows of each table show the MA-GIRFs for the appropriate time sample. In each case, we find that the multiplier in times of slack is slightly larger than in normal times but not significantly so. Moreover, in neither sample are the MA-GIRF multipliers larger than 1.

4.3.2 Transition Variable

Our baseline results use the output gap as the transition variable to differentiate periods of slack and nonslack. In this subsection, we check for robustness using the unemployment rate as alternative transition variable and a baseline VAR(4). The last two columns of Table 3 and Table 4 contain these results for the full sample and the short sample, respectively. The most direct comparison is the third and fourth columns that use the same baseline VAR(4) but with the output gap. Here, again, we report results using the BP identification; results obtained using the Ramey news shocks are in the Appendix.

For the full sample, the estimated multipliers are less than one and are only significantly difference across business cycle regimes for RD-LP. For the shorter sample, using the unemployment rate does appear to affect the point estimates of the regime-dependent multipliers. For the MSVAR, the multipliers are larger; for the STVAR, the multipliers are smaller. For RD-LP, the point values of the multipliers are nearly identical.

How do we resolve the apparent differences across the transition variables? We compute the MA-GIRFs including all models with both transition variables. Here, we find results more consistent with our baseline results. For the full sample, the multipliers are small (on the order of 0.4) and not statistically significant across regimes. For the shorter sample, the multipliers are relatively larger than in the full sample (around 0.8 in times of slack and 0.4 in normal times) and have larger error bands.

4.3.3 Comprehensive MA Multiplier

Given the copious model and identification uncertainty in the literature, we compute a comprehensive model average multiplier derived from variations of the four model specifications discussed above.[18] We estimate models varying (i) the transition process (TVAR, MSVAR, or STVAR), (ii) the shock identification (Ramey News or Blanchard-Perotti), (ii) the VAR lag length (one, four, or six), and (iv) the transition variable (output gap or unemployment rate).

We estimate the regime-dependent multipliers averaging over combinations of the four variations above when using, alternately, the full and short samples.[19] For the full sample, the multiplier in the slack and no-slack regimes are 0.64 and 0.54, respectively. In both regimes, the multipliers are statistically below unity. Consistent with the narrative espoused by some papers in the literature, the point estimates of the multiplier in the slack regime is larger but not statistically different from the non-slack regime. In the shorter sample, the multipliers are closer to zero (0.40 in slack and 0.26 in non-slack), still seemingly below unity, and, again, not importantly different from each other. The comprehensive multipliers across self-exciting VARs reinforce the findings of RZ.

In summary, the comprehensive multiplier finds no significant difference across states whereas RD-LP finds significant differences when using BP identification. Our simulation

[18] As we noted above, the weights in the MA are applied directly to the IRFs. Because the object of interest is the multiplier, we could apply the weights directly to the multiplier. The appendix presents the estimated multipliers when using this approach and they are quantitatively the same as when averaging over the underlying IRFs.

[19] Error bands for the comprehensive multipliers are presented in the online appendix.

results suggest the RD-LP estimates will be more biased as long as a sufficiently long lag length is used in the VAR specifications. Together with the findings of Gonçalves et al. (2023), care should be taken when considering estimates from regime-dependent RD-LP.

5 Conclusion

We compare the accuracy of GIRFs in self-exciting nonlinear models in the presence of model uncertainty and other possible misspecifications. In particular, we evaluate, using MC simulation experiments, how these misspecifications affect the ability to approximate the true IRF and capture state asymmetry. In addition, we introduce a MA-GIRF that weights the IRFs of based on the BIC.

We arrive at a few main conclusions based on our MC experiments. If one has a strong model prior about the true form of the underlying state process, estimating that model is typically first best. Any significant uncertainty about model specification, however, can be ameliorated by estimating an MA-GIRF over a model set that includes VARs with sufficient lag order to capture the relevant dynamics. In the case that the appropriate VAR is suspected to have an inestimable number of parameters, estimating regime-dependent local projections is a viable alternative with its own caveats. Our results complement those in Gonçalves et al. (2023), who explore the effect of variation in the shock size in similar models. Methods used here can be augmented with those in Gonçalves et al. (2023), providing a comprehensive approach to modeling IRFs in the presence of nonlinearities and model uncertainty.

Based on these findings, we address the problem of state-dependency in the fiscal spending multiplier. We estimate a variety of nonlinear VARs to obtain GIRFs and the MA-GIRF. We find that fiscal multipliers are consistently below unity, irrespective of the shock, the transition variable, or number of VAR lags used in estimation, when estimated over a long sample dating back to 1890. However, the regime-dependent multipliers are sensitive to the use of only post-Korean war data.[20] Therefore, studies, using postwar data, which find significant differences in fiscal multipliers across business cycle regimes will likely not be robust to model misspecification or alternative shock identifications.

[20]The multiplier may also be sensitive to the source of the economic shock. Ghassibe and Zanetti (2022) propose a model where the effectiveness of fiscal policy depends on whether the policy stimulates supply or demand and whether the source of economic fluctuations is supply or demand shocks. They find that matching the fiscal policy to the source of the shock leads to higher fiscal multipliers.

References

Alloza, Mario (2022) "Is Fiscal Policy More Effective During Recessions?" *International Economic Review*, 63 (3), 1271–1292, https://doi.org/10.1111/iere.12571.

Auerbach, Alan J. and Yuriy Gorodnichenko (2012) "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4 (2), 1–27, 10.1257/pol.4.2.1.

Auerbach, Alan J and Yuriy Gorodnichenko (2013) "Fiscal multipliers in recession and expansion," in *Fiscal policy after the financial crisis*, 63–98: University of Chicago Press.

Balke, Nathan (2000) "Credit and Economic Activity: Credit Regimes and Nonlinear Propagation of Shocks," *The Review of Economics and Statistics*, 82 (2), 344–349, `https://EconPapers.repec.org/RePEc:tpr:restat:v:82:y:2000:i:2:p:344-349`.

Barnichon, Regis, Davide Debortoli, and Christian Matthes (2021) "Understanding the Size of the Government Spending Multiplier: It's in the Sign," *The Review of Economic Studies*, 89 (1), 87–117, 10.1093/restud/rdab029.

Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K. Van Dijk (2016) "Interconnections Between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model," *Journal of Applied Econometrics*, 31 (7), 1352–1370, https://doi.org/10.1002/jae.2501.

Blanchard, Olivier and Roberto Perotti (2002) "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *The Quarterly Journal of Economics*, 117 (4), 1329–1368, `http://www.jstor.org/stable/4132480`.

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari (2015) "Estimating Fiscal Multipliers: News from a Non-linear World," *The Economic Journal*, 125 (584), 746–776, 10.1111/ecoj.12263.

Castelnuovo, Efrem and Guay Lim (2019) "What Do We Know About the Macroeconomic Effects of Fiscal Policy? A Brief Survey of the Literature on Fiscal Multipliers," *Australian Economic Review*, 52 (1), 78–93, https://doi.org/10.1111/1467-8462.12313.

Diebold, Francis, Joon-Haeng Lee, and Gretchen Weinbach (1993) "Regime switching with time-varying transition probabilities," Working Papers 93-12, Federal Reserve Bank of Philadelphia, `https://EconPapers.repec.org/RePEc:fip:fedpwp:93-12`.

Fazzari, Steven M., James Morley, and Irina Panovska (2015) "State-dependent effects of fiscal policy," *Studies in Nonlinear Dynamics & Econometrics*, 19 (3), 285–315, doi:10.1515/snde-2014-0022.

——— (2021) "When is discretionary fiscal policy effective?" *Studies in Nonlinear Dynamics & Econometrics*, 25 (4), 229–254, doi:10.1515/snde-2018-0113.

Ferraresi, Tommaso, Andrea Roventini, and Giorgio Fagiolo (2015) "Fiscal Policies and Credit Regimes: A TVAR Approach," *Journal of Applied Econometrics*, 30 (7), 1047–1072, https://doi.org/10.1002/jae.2420.

Filardo, Andrew J. (1994) "Business-Cycle Phases and Their Transitional Dynamics," *Journal of Business & Economic Statistics*, 12 (3), 299–308, 10.1080/07350015.1994.10524545.

Fisher, Jonas D.M. and Ryan Peters (2010) "Using Stock Returns to Identify Government Spending Shocks," *The Economic Journal*, 120 (544), 414–436, `http://www.jstor.org/stable/27765780`.

Gechert, Sebastian and Ansgar Rannenberg (2018) "Which Fiscal Multipliers Are Regime-Dependent? A Meta-Regression Analysis," *Journal of Economic Surveys*, 32 (4), 1160–1182, https://doi.org/10.1111/joes.12241.

Geweke, John and Gianni Amisano (2011) "Optimal prediction pools," *Journal of Econometrics*, 164 (1), 130–141.

Ghassibe, Mishel and Francesco Zanetti (2022) "State dependence of fiscal multipliers: the source of fluctuations matters," *Journal of Monetary Economics*, 132, 1–23, https://www.sciencedirect.com/science/article/pii/S0304393222001155.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento (2023) "State-Dependent Local Projections," Working Paper 2302, Federal Reserve Bank of Dallas.

Gordon, Robert J and Robert Krenn (2010) "The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers," Working Paper 16380, National Bureau of Economic Research, 10.3386/w16380.

Granger, Clive and Timo Teräsvirta (1993) *Modelling Non-Linear Economic Relationships*: Oxford University Press, https://EconPapers.repec.org/RePEc:oxp:obooks:9780198773207.

Hansen, Bruce E. (2007) "Least Squares Model Averaging," *Econometrica*, 75 (4), 1175–1189, https://doi.org/10.1111/j.1468-0262.2007.00785.x.

Hansen, Bruce E (2016) "Stein combination shrinkage for vector autoregressions," *Manuscript, University of Wisconsin-Madison*.

Ho, Paul, Thomas A. Lubik, and Christian Matthes (2023) "Averaging Impulse Responses Using Prediction Pools," *Federal Reserve Bank of Richmond Working Paper 23-04*, https://doi.org/10.21144/wp23-04.

Inoue, Atsushi, Barbara Rossi, and Yiru Wang (2022) "Local Projections in Unstable Environments: How Effective is Fiscal Policy?".

Jordà, Òscar (2005) "Estimation and inference of impulse responses by local projections," *American economic review*, 95 (1), 161–182.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95 (1), 161–182, 10.1257/0002828053828518.

Jørgensen, Peter L. and Søren H. Ravn (2022) "The inflation response to government spending shocks: A fiscal price puzzle?" *European Economic Review*, 141, 103982, https://doi.org/10.1016/j.euroecorev.2021.103982.

Kang, Jihye and Soyoung Kim (2022) "Government spending news and surprise shocks: It's the timing and persistence," *Journal of Macroeconomics*, 73, 103446, https://doi.org/10.1016/j.jmacro.2022.103446.

Kilian, Lutz and Yun Jung Kim (2011) "How Reliable Are Local Projection Estimators of Impulse Responses?," *The Review of Economics and Statistics*, 93 (4), 1460–1466, 10.1162/REST_a_00143.

Koop, Gary, M.Hashem Pesaran, and Simon M. Potter (1996) "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74 (1), 119–147, https://doi.org/10.1016/0304-4076(95)01753-4.

Krolzig, Hans-Martin (2006) "Impulse-Response Analysis in Markov Switching Vector Autoregressive Models," Mimeo, University of Kent.

Krüger, Fabian, Sebastian Lerch, Thordis L. Thorarinsdottir, and Tilmann Gneiting (2016) "Predictive Inference Based on Markov Chain Monte Carlo Output," 10.48550/ARXIV.1608.06802.

Laumer, Sebastian and Collin S. Phillips (2020) "Does the Government Spending Multiplier Depend on the Business Cycle?" *Journal of Money, Credit, and Banking*, forthcoming.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf (2021) "Local Projections vs. VARs: Lessons From Thousands of DGPs," 10.48550/ARXIV.2104.00655.

Lusompa, Amaze (2021) "Local projections, autocorrelation, and efficiency," *Federal Reserve Bank of Kansas City Working Paper* (21-01).

Matheson, James E. and Robert L. Winkler (1976) "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22 (10), 1087–1096, 10.1287/mnsc.22.10.1087.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2021) "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89 (4), 1789–1823, https://doi.org/10.3982/ECTA18756.

Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf (2024) "Double Robustness of Local Projections and Some Unpleasant VARithmetic," Working Paper 32495, National Bureau of Economic Research, http://www.nber.org/papers/w32495.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy (2013) "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data," *American Economic Review*, 103 (3), 129–34, 10.1257/aer.103.3.129.

Plagborg-Møller, Mikkel and Christian K Wolf (2021) "Local projections and VARs estimate the same impulse responses," *Econometrica*, 89 (2), 955–980.

Raftery, Adrian E. (1995) "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111–163, http://www.jstor.org/stable/271063.

Ramey, Valerie A. (2011) "Identifying Government Spending Shocks: It's all in the Timing*," *The Quarterly Journal of Economics*, 126 (1), 1–50, 10.1093/qje/qjq008.

——— (2016) "Macroeconomic Shocks and Their Propagation," Chapter 2 of Handbook of Macroeconomics, 71–162: Elsevier, https://doi.org/10.1016/bs.hesmac.2016.03.003.

——— (2019) "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives*, 33 (2), 89–114, 10.1257/jep.33.2.89.

Ramey, Valerie A. and Sarah Zubairy (2018) "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126 (2), 850–901, 10.1086/696277.

Ruisi, Germano (2019) "Time-varying local projections," Working Paper 891, London, http://hdl.handle.net/10419/210448.

Steel, Mark F. J. (2020) "Model Averaging and Its Use in Economics," *Journal of Economic Literature*, 58 (3), 644–719, 10.1257/jel.20191385.

Swanson, Norman R. and Tian Zeng (2001) "Choosing among competing econometric forecasts: Regression-based forecast combination using model selection," *Journal of Forecasting*, 20 (6), 425–440, https://doi.org/10.1002/for.784.

Table 1: **CRPS Across Simulation Specifications**

This table shows the mean CRPS for each estimation method for a given true data generating process and specification error. We compute the overall CRPS as the average over 20 impulse response horizons. The 68% posterior interval is shown in brackets beneath the mean.

Specification Error:	**True Model:**	**Estimated Model:**				
		TVAR	*MSVAR*	*STVAR*	*RD-LP*	*MA*
Transition	TVAR	0.20 [0.09,0.55]	0.42 [0.17,0.75]	0.66 [0.40,1.03]	0.66 [0.33,1.15]	0.24 [0.12,0.56]
Process	MSVAR	0.16 [0.09,0.35]	0.12 [0.06,0.31]	0.21 [0.12,0.41]	0.20 [0.13,0.37]	0.14 [0.08,0.30]
Only	STVAR	0.12 [0.06,0.27]	0.13 [0.06,0.28]	0.14 [0.08,0.28]	0.25 [0.15,0.55]	0.12 [0.06,0.25]
Incorrect Lag	TVAR	0.38 [0.20,0.80]	0.49 [0.22,1.31]	0.71 [0.39,1.45]	0.53 [0.30,1.06]	0.44 [0.22,1.00]
on Transition	MSVAR	0.15 [0.09,0.29]	0.12 [0.06,0.23]	0.16 [0.11,0.28]	0.20 [0.13,0.33]	0.12 [0.07,0.22]
Variable	STVAR	0.13 [0.07,0.23]	0.10 [0.05,0.22]	0.13 [0.07,0.24]	0.28 [0.14,0.59]	0.11 [0.06,0.21]
Incorrect	TVAR	0.42 [0.23,0.82]	0.49 [0.23,1.28]	0.77 [0.42,1.40]	0.57 [0.33,1.09]	0.47 [0.24,1.05]
Transition	MSVAR	0.17 [0.10,0.33]	0.14 [0.07,0.28]	0.21 [0.13,0.35]	0.20 [0.14,0.34]	0.14 [0.08,0.28]
Variable	STVAR	0.18 [0.09,0.34]	0.11 [0.06,0.27]	0.13 [0.08,0.34]	0.31 [0.16,0.54]	0.12 [0.06,0.28]
Incorrect	TVAR	1.10 [0.73,1.91]	1.25 [0.84,2.08]	1.51 [0.93,2.34]	0.60 [0.36,1.18]	1.12 [0.76,1.95]
VAR	MSVAR	0.63 [0.51,0.75]	0.63 [0.53,0.78]	0.65 [0.55,0.81]	0.21 [0.14,0.33]	0.60 [0.49,0.71]
Lag (1)	STVAR	0.25 [0.08,0.49]	0.26 [0.09,0.49]	0.23 [0.09,0.51]	0.34 [0.16,0.58]	0.25 [0.08,0.49]
Incorrect	TVAR	0.25 [0.10,0.69]	0.45 [0.16,1.18]	0.69 [0.38,1.41]	0.54 [0.31,1.07]	0.31 [0.13,0.72]
VAR	MSVAR	0.28 [0.16,0.54]	0.22 [0.14,0.42]	0.42 [0.28,0.60]	1.31 [0.23,4.23]	0.26 [0.15,0.45]
Lag (6)	STVAR	0.16 [0.08,0.32]	0.17 [0.09,0.33]	0.14 [0.09,0.29]	0.35 [0.18,0.67]	0.15 [0.08,0.28]
True Model is Linear	VAR	1.67 [1.14,2.49]	1.36 [0.78,2.09]	1.17 [0.56,1.84]	1.54 [0.85,2.05]	1.49 [0.96,2.21]

Table 2: **Fiscal Spending Multipliers**

This table shows the median posterior five-year fiscal multiplier for each model under each business cycle regime. The 68% highest posterior density interval is shown below in brackets. The multiplier is computed as the cumulative response of output divided by the cumulative response of government spending. The first and second columns show the multiplier based on the Ramey news shock series under slack and nonslack regimes, respectively. The third and fourth columns show similar estimates when using Blanchard-Perotti shock identification. In each case the shock size is a one-percent of GDP increase in government spending. Four lags of Y_t are used in both the VAR and RD-LP specifications. The last row shows the multiplier implied by the model average impulse response when averaging over all three model specifications (TVAR, MSVAR, and STVAR) as well as both shocks (Ramey News and Blanchard-Perotti).

(a) Full Sample

Sample:	1890Q1-2015Q4			
Shock ID:	Ramey News		Blanchard-Perotti	
	Slack	Nonslack	Slack	Nonslack
TVAR	0.49 [0.15,0.87]	0.59 [0.37,0.83]	0.42 [0.13,0.68]	0.34 [0.18,0.51]
MSVAR	0.93 [0.57,1.34]	0.67 [0.33,1.00]	0.59 [0.45,0.73]	0.32 [0.13,0.50]
STVAR	0.98 [0.45,1.54]	0.90 [0.35,1.41]	0.49 [0.35,0.66]	0.29 [0.20,0.40]
Model Avg	0.53 [0.17,0.92]	0.59 [0.38,0.83]	0.46 [0.23,0.65]	0.31 [0.18,0.46]
RD-LP	0.53 [0.46,0.61]	0.68 [0.55,0.81]	0.79 [0.72,0.87]	0.31 [0.24,0.37]
		Slack	Nonslack	
Model Avg (Both Shock IDs)		0.46 [0.24,0.66]	0.33 [0.20,0.50]	

(b) Short Sample

Sample:	1969Q1-2015Q4			
Shock ID:	Ramey News		Blanchard-Perotti	
	Slack	Nonslack	Slack	Nonslack
TVAR	1.49 [−3.68,7.44]	1.88 [−7.94,11.37]	0.83 [−0.05,1.71]	0.02 [−1.19,1.12]
MSVAR	0.67 [−8.33,11.00]	0.97 [−5.62,7.74]	0.95 [−0.11,2.07]	1.06 [−0.01,2.20]
STVAR	0.97 [−3.93,5.72]	2.26 [−11.25,14.68]	1.53 [0.68,2.45]	0.33 [−0.55,1.11]
Model Avg	1.27 [−4.24,7.37]	1.65 [−8.39,11.63]	0.84 [0.02,1.71]	0.30 [−0.85,1.37]
RD-LP	−0.24 [−0.95,0.48]	−2.00 [−8.15,2.68]	0.56 [0.34,0.78]	−2.59 [−4.04,−1.49]
		Slack	Nonslack	
Model Avg (Both Shock IDs)		0.83 [−0.13,1.80]	0.27 [−1.09,1.50]	

Table 3: **Fiscal Spending Multipliers – Robustness Checks for Full Sample**

This table shows the median posterior five-year fiscal multiplier for each model under each business cycle regime when using the full sample. The first six columns show the multiplier estimates across different VAR lags when using the output gap as the transition variable. The last two columns show similar estimates when using four VAR lags and the unemployment rate as the transition variable. The last two rows of each panel show the multiplier implied by the model average impulse response when averaging over all three model specifications (TVAR, MSVAR, and STVAR) as well as either the VAR lags (second to last row) or the two transition variables (last row). The 68% highest posterior density interval is shown below in brackets.

Sample:	1890Q1-2015Q4							
Shock ID:	Blanchard-Perotti							
Transition Var:	Output Gap		Output Gap		Output Gap		UR	
VAR Lags	1		4		6		4	
	Slack	Nonslack	Slack	Nonslack	Slack	Nonslack	Slack	Nonslack
TVAR	0.47	0.44	0.42	0.34	0.58	0.47	0.65	0.44
	[0.28,0.88]	[0.34,0.55]	[0.13,0.68]	[0.18,0.51]	[0.34,0.79]	[0.33,0.61]	[0.40,0.89]	[0.28,0.60]
MSVAR	0.01	0.51	0.59	0.32	0.54	0.56	0.27	−0.02
	[−0.37,0.36]	[0.36,0.65]	[0.45,0.73]	[0.13,0.50]	[0.32,0.74]	[0.38,0.73]	[0.08,0.44]	[−0.39,0.25]
STVAR	0.40	0.28	0.49	0.29	0.72	0.44	0.67	0.49
	[0.30,0.63]	[0.23,0.33]	[0.35,0.66]	[0.20,0.40]	[0.49,1.08]	[0.33,0.55]	[0.57,0.78]	[0.40,0.58]
Model Avg	0.42	0.39	0.46	0.31	0.55	0.48	0.65	0.44
	[0.28,0.75]	[0.28,0.52]	[0.23,0.65]	[0.18,0.46]	[0.37,0.74]	[0.34,0.63]	[0.40,0.89]	[0.28,0.60]
RD-LP	0.44	0.69	0.79	0.31	0.28	0.52	0.69	0.25
	[0.38,0.50]	[0.64,0.74]	[0.72,0.87]	[0.24,0.38]	[0.19,0.36]	[0.45,0.58]	[0.61,0.77]	[0.18,0.32]
			Slack	Nonslack				
Model Avg (All VAR Lags)			0.43	0.37				
			[0.30,0.65]	[0.27,0.49]				
Model Avg (Both Transition Variables)			0.45	0.33				
			[0.24,0.64]	[0.19,0.50]				

Table 4: **Fiscal Spending Multipliers – Robustness Checks for Short Sample**

This table shows the median posterior five-year fiscal multiplier for each model under each business cycle regime when using the short sample. The first six columns show the multiplier estimates across different VAR lags when using the output gap as the transition variable. The last two columns show similar estimates when using four VAR lags and the unemployment rate as the transition variable. The last two rows of each panel show the multiplier implied by the model average impulse response when averaging over all three model specifications (TVAR, MSVAR, and STVAR) as well as either the VAR lags (second to last row) or the two transition variables (last row). The 68% highest posterior density interval is shown below in brackets.

Sample:	1969Q1-2015Q4							
Shock ID:	Blanchard-Perotti							
Transition Var:	Output Gap		Output Gap		Output Gap		UR	
VAR Lags	1		4		6		4	
	Slack	Nonslack	Slack	Nonslack	Slack	Nonslack	Slack	Nonslack
TVAR	0.52 [0.10,0.95]	0.30 [−0.12,0.71]	0.83 [−0.05,1.71]	0.02 [−1.19,1.12]	0.83 [−0.16,1.87]	0.63 [−0.47,1.92]	0.33 [−0.53,1.09]	0.44 [−0.68,1.47]
MSVAR	0.18 [−0.45,0.79]	0.28 [−0.33,0.88]	0.95 [−0.11,2.07]	1.06 [−0.01,2.20]	0.08 [−1.03,1.09]	0.10 [−1.01,1.09]	1.49 [0.77,2.18]	1.70 [0.94,2.40]
STVAR	0.70 [−0.03,1.37]	−0.29 [−0.92,0.22]	1.53 [0.68,2.45]	0.33 [−0.55,1.11]	1.41 [0.36,2.43]	0.27 [−0.83,1.37]	0.06 [−0.83,0.85]	−0.18 [−1.19,0.69]
Model Avg	0.36 [−0.13,0.79]	0.26 [−0.16,0.67]	0.84 [0.02,1.71]	0.30 [−0.85,1.37]	0.64 [−0.34,1.63]	0.49 [−0.56,1.62]	0.80 [−0.14,1.69]	1.00 [−0.20,1.99]
RD-LP	0.41 [0.22,0.62]	−1.60 [−2.39,−0.92]	0.56 [0.33,0.79]	−2.56 [−4.07,−1.49]	0.53 [0.28,0.77]	−1.60 [−2.96,−0.52]	0.58 [0.37,0.79]	−2.02 [−3.34,−1.04]
			Slack	Nonslack				
Model Avg (All VAR Lags)			0.40 [−0.09,0.83]	0.26 [−0.18,0.69]				
Model Avg (Both Transition Variables)			0.87 [0.04,1.74]	0.42 [−0.68,1.43]				

Figure 1: **CRPS By Horizon**

This figure shows the median CRPS for each impulse response horizon for each estimation method. The data generating process for the top, middle, and bottom panels are a TVAR, MSVAR, and STVAR, respectively.

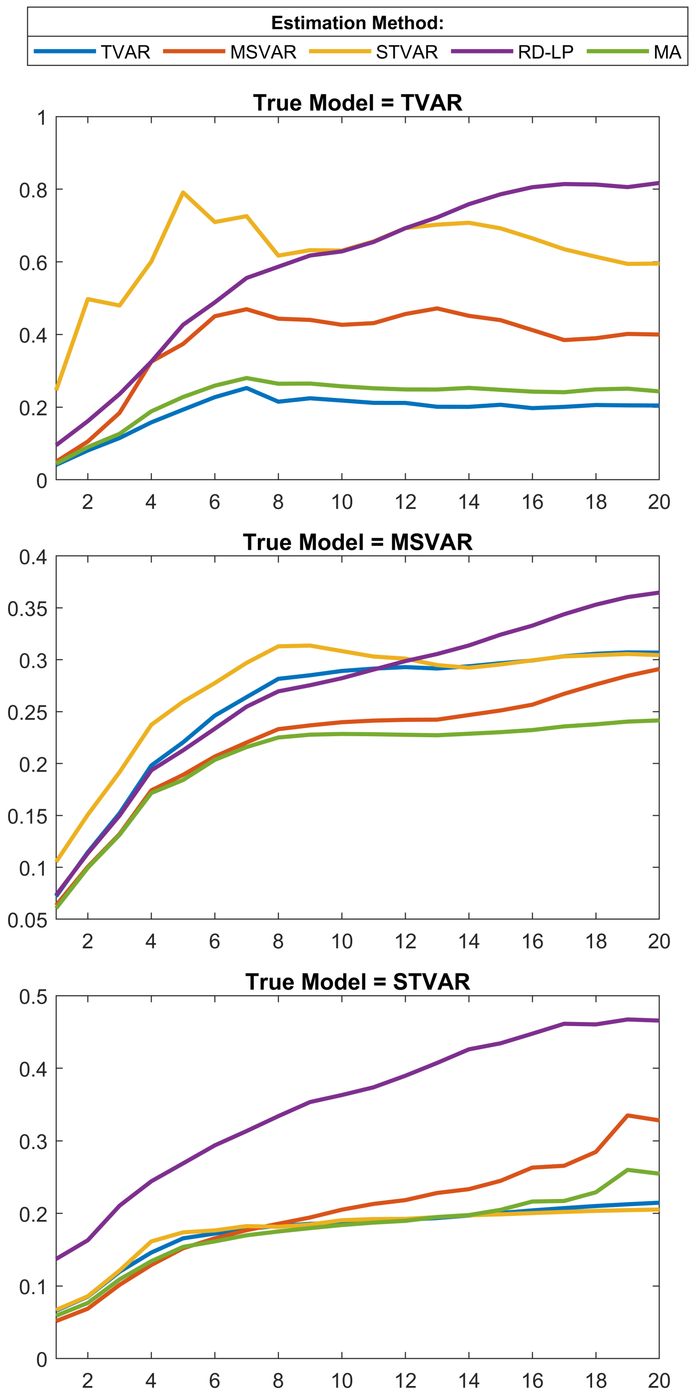

Figure 2: **Histogram of Reported Regime-Dependent Fiscal Multipliers**.

This histogram shows the reported regime-dependent multipliers from the meta-analysis conducted by Gechert and Rannenberg (2018) which covers 98 previous studies. The blue bars show the estimated fiscal multipliers in normal economic regimes and the red bars show the multipliers in slack regimes. The solid vertical lines show the average multiplier in each regime across the previous studies.

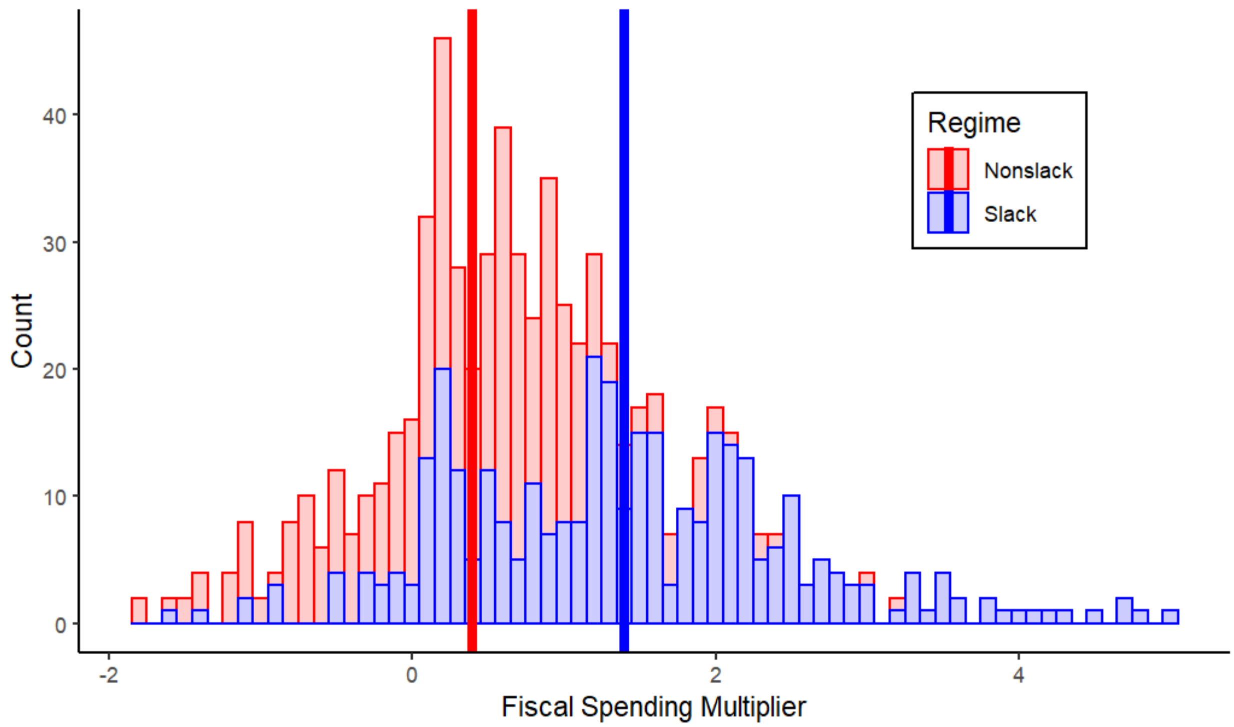

Figure 3: **Model Average GIRFs**

This figure shows the model average impulse responses for government spending and output when using the Ramey news shock series for the full sample (1890Q1-2015Q4). For each variable, the model average weighs the GIRF from the TVAR, MSVAR, and STVAR based on their respective BIC. Four lags of Y_t are used in each VAR specification.