Joint Bayesian Inference about Impulse Responses in VAR Models

Atsushi Inoue*
Vanderbilt University

Lutz Kilian†
Federal Reserve Bank of Dallas
CEPR

First draft: January 21, 2020. This version: November 2, 2020.

Abstract: We derive the Bayes estimator of vectors of structural VAR impulse responses under a range of alternative loss functions. We also derive joint credible regions for vectors of impulse responses as the lowest posterior risk region under the same loss functions. We show that conventional impulse response estimators such as the posterior median response function or the posterior mean response function are not in general the Bayes estimator of the impulse response vector obtained by stacking the impulse responses of interest. We show that such pointwise estimators may imply response function shapes that are incompatible with any possible parameterization of the underlying model. Moreover, conventional pointwise quantile error bands are not a valid measure of the estimation uncertainty about the impulse response vector because they ignore the mutual dependence of the responses. In practice, they tend to understate substantially the estimation uncertainty about the impulse response vector.

JEL code: C22, C32, C52

Key words: Loss function, joint inference, median response function, mean response function, modal model, posterior risk.

*Vanderbilt University, Department of Economics, Nashville, TN 37235-1819. E-mail: atsushi.inoue@vanderbilt.edu.

†Federal Reserve Bank of Dallas, Research Department, 2200 N. Pearl St., Dallas, TX 75201, USA. E-mail: lkilian2019@gmail.com (corresponding author).

1 Introduction

It is standard practice in empirical macroeconomics to report estimates of impulse response functions based on structural vector autoregressive (VAR) models (see Kilian and Lütkepohl 2017). Applied users are typically interested in the shape of the impulse response functions. Key questions include whether the responses are hump shaped, delayed or sluggish, for example (e.g., Cochrane 1994; Woodford 2003; Eichenbaum and Evans 1995). Applied users are also typically interested in the comovement of the responses across model variables (e.g., Blanchard 1989; Sims and Zha 1999; Inoue and Kilian 2016). These questions cannot be answered by studying one impulse response at a time, but require inference about the vector θ of impulse responses obtained by stacking all impulse responses of interest.

In this paper, we provide a unified approach to the evaluation of vectors of structural impulse responses from a Bayesian point of view. Our approach differs conceptually and practically from the established practice in this literature. Our first main contribution is to use Bayesian decision theory to formally derive estimators of the impulse response vector θ under commonly used loss functions including absolute loss, quadratic loss and Dirac delta loss, building on Bernardo (2010). These estimators take the form $\widehat{\overline{\theta}} = \arg\min E_\theta[L(\theta, \overline{\theta})]$, where $\overline{\theta}$ denotes an action and $L(\theta, \overline{\theta})$ the loss function. Such estimators are new to the VAR literature and can be easily implemented in typical applications. We also show how these estimators can be made more robust to changes in the scaling of the data and of the impulse responses, and we discuss how to address the problem of the potential nonuniqueness of the estimator.

We demonstrate that conventional estimators of θ such as the posterior median response function, defined as the vector of medians obtained from the marginal posterior distribution of each individual impulse response, as well as analogously defined posterior mean response functions are not in general the Bayes estimator of θ under additively separable absolute and quadratic loss,

respectively. Only in the special case when the number of structural impulse responses (n_{irf}) does not exceed the number of structural model parameters (n_p) do they coincide with the Bayes estimator in the limit, as the number of posterior draws approaches infinity. This result follows from the fact that, for any additively separable loss function,

$$\begin{aligned} E_\theta[L(\theta,\bar{\theta})] &= \int L_1(\theta_1,\bar{\theta}_1)f(\theta)d\theta + \cdots \int L_{n_{irf}}(\theta_{n_{irf}},\bar{\theta}_{n_{irf}})f(\theta)d\theta \\ &= \int L_1(\theta_1,\bar{\theta}_1)f_1(\theta_1)d\theta_1 + \cdots \int L_{n_{irf}}(\theta_{n_{irf}},\bar{\theta}_{n_{irf}})f_{n_{irf}}(\theta_{n_{irf}})d\theta_{n_{irf}} \\ &= E_{\theta_1}[L_1(\theta_1,\bar{\theta}_1)] + \cdots + E_{\theta n_{irf}}[L_{n_{irf}}(\theta_{n_{irf}},\bar{\theta}_{n_{irf}})] \end{aligned}$$

where $f(\cdot)$ denotes the joint density of θ, $f_i(\cdot)$, $i = 1, ..., n_{irf}$, denotes the marginal density with respect to the ith element of θ, and $n_{irf} \leq n_p$ is used in the second equality.

In contrast, when $n_{irf} > n_p$, as is the case in typical empirical applications, posterior median response functions and mean response functions even in the limit are not the Bayes estimator of θ. We provide analytical examples that illustrate that posterior median and mean response functions in this setting may imply impulse response function shapes that differ from the shape of the response function under any possible parameterization of the underlying autoregressive model. Our analysis articulates more precisely concerns first raised in Fry and Pagan (2011), Kilian and Murphy (2012) and Inoue and Kilian (2013) that median response functions tend to conflate the dynamic responses implied by different structural models as well as their comovement, calling into question the economic interpretation of the posterior median response function.

We conclude that the practice of reporting posterior median (or mean) response functions in applied work, which has been standard in the Bayesian VAR literature for the last four decades, is inconsistent with the objectives of most applied users who are interested in the dynamics of the responses across horizons and across model variables and may be misleading in practice. Our

results also show that Baumeister and Hamilton's (2018) claim that the posterior median response function is optimal from a decision theoretic point of view does not apply to θ in general. This is not to dispute that posterior median (and mean) response functions are valid summary statistics for the central tendency of the marginal impulse response posterior distributions, but this result is of limited use because there are few (if any) examples of economically interesting questions that can be answered based on these statistics.

Our second main contribution is to derive estimators of the joint credible set under the same loss function used in deriving the estimator of θ, building on the work of Bernardo (2010). It is well known that conventional Bayesian error bands, defined as vectors of upper and lower quantiles of the marginal impulse response posterior distributions misrepresent the estimation uncertainty about estimates of impulse response vectors. This point is neither new nor controversial.[1] It is well documented in the Bayesian literature, and it applies to both point identified and set-identified models (e.g., Sims and Zha 1999; Inoue and Kilian 2013).[2] The only reason that the use of pointwise Bayesian error bands has persisted in applied work has been the lack of an easy to implement alternative that can be applied to a wide range of structural VAR models. In this paper, we propose a new approach to the construction of joint credible sets for structural impulse responses that can be adapted to any of the loss functions of interest.[3]

Our joint credible set is defined as the lowest posterior risk region $\left\{\overline{\theta} \in \Theta : E_\theta[L(\theta, \overline{\theta})] \leq c_{1-\alpha,L}\right\}$, where $c_{1-\alpha,L}$ denotes the smallest number such that the posterior probability of the credible set

[1] As noted by Sims and Zha (1999), bands constructed by "connecting the dots" tend to be misleading because it is extremely unlikely in economic applications that the uncertainty about the impulse response estimates is independent across horizons. Moreover, as emphasized in Inoue and Kilian (2016) and Kilian and Lütkepohl (2017), the dependence across impulse responses matters not only for conducting joint inference across a range of horizons for a given impulse response function, but it matters equally for conducting joint inference about multiple impulse response functions.

[2] Analogous concerns have been raised in the growing frequentist literature on simultaneous confidence regions for impulse responses (e.g., Lütkepohl, Staszewska-Bystrova and Winker 2015a,b, 2018; Inoue and Kilian 2016; Bruder and Wolf 2018; Montiel Olea and Plagborg-Møller 2019).

[3] In related work, Montiel Olea and Plagborg-Møller (2019) propose a simultaneous credible band for impulse response functions based on the product of component-wise equal-tailed credible intervals, where the tail probability has been calibrated to yield simultaneous credibility $1 - \alpha$.

is $1-\alpha$. These credible sets are guaranteed to generate solutions contained in the set of impulse response functions that can be generated by the underlying autoregressive model. They not only properly account for the estimation uncertainty about the vector θ, which may be several times larger than implied by pointwise methods of inference, but, unlike joint error bands, they convey additional information about the directions in which the estimates are likely to depart from the estimate. This proposal directly addresses the concern raised in Sims and Zha (1999, p. 1113) that conventional error bands do not provide much information about the deviations from the estimate of the response function that are most likely. We show that these joint credible sets may be graphically represented analogous to shot-gun trajectory plots in ballistics, preserving information about likely departures from the estimated path of θ that is discarded when constructing joint error bands. Different colors may be used to highlight information contained in these credible sets about the shape of the response functions and their comovement.

The remainder of the paper is organized as follows. In section 2, we discuss estimation and joint inference under absolute and quadratic loss. We also compare our approach to pointwise inference under absolute and quadratic loss, and we illustrate why the conventional approach in the literature can be misleading. The key difference between our approach and conventional median and mean response functions can be summarized as follows. The impulse response vector, $\theta = g(\lambda)$, is a vector-valued function that maps from the space $\Lambda \subset \Re^{n_p}$ of the underlying structural VAR parameters, $\lambda \in \Lambda$, to the space $\Theta \subset \Re^{n_{irf}}$ of the structural responses. Note that this function is surjective on the co-domain $\Theta = g(\Lambda)$, but not on its complement Θ^c.[4] For concreteness, suppose that the loss function is additively quadratic and denote the data by Y. Then the conventional optimal Bayes decision, $\tilde{\tilde{\theta}} = [E_{\theta_1}(\theta_1|Y)\ E_{\theta_2}(\theta_2|Y)\ \cdots\ E_{\theta_{n_{irf}}}(\theta_{n_{irf}}|Y)]'$ for pointwise inference, known as the mean response function, minimizes $E_{\theta_1}[(\bar{\theta}_1 - \theta_1)^2|Y] + E_{\theta_2}[(\bar{\theta}_2 - \theta_2)^2|Y] + \cdots +$

[4] A function g from a set Λ to a set Θ is surjective, if for every element θ in the codomain Θ of g, there is at least one element λ in the domain Λ of g such that $g(\lambda) = \theta$. It is not required that λ be unique; the function g may map one or more elements of Λ to the same element of Θ.

$E_{\theta_{nirf}}[(\bar{\theta}_{n_{irf}} - \theta_{n_{irf}})^2|Y]$ for all $\bar{\theta} \in \Re^{n_{irf}}$, where the expectation is taken with respect to the maginal posterior distributions of θ. In contrast, the optimal Bayes decision proposed in this paper, denoted $\widehat{\overline{\theta}}$, minimizes $E_\theta\left[\left\|\overline{\theta} - \theta\right\|^2 |Y\right]$, where the expectation is taken with respect to the joint posterior distribution of θ and $\|\cdot\|$ denotes the Euclidian norm, among all $\overline{\theta} \in \Theta$, for which there exists a λ in Λ such that $g(\lambda) = \delta$. Although the loss function in our analysis is the same as in the conventional approach, the constraint set is different. By construction, our optimal solution $\widehat{\overline{\theta}}$ always belongs to Θ, so there always exists a $\widehat{\overline{\lambda}} \in \Lambda$ such that $\widehat{\overline{\theta}} = g(\widehat{\overline{\lambda}})$. This is not the case for the conventional estimator. If $\widetilde{\overline{\theta}}$ belongs to Θ, there exists $\widetilde{\overline{\lambda}} \in \Lambda$ such that $\widetilde{\overline{\theta}} = g(\widetilde{\overline{\lambda}}) = E_\theta(\theta|Y)$ because g is surjective on Θ. In contrast, if $\widetilde{\overline{\theta}}$ belongs to Θ^c, $\widehat{\overline{\theta}}$ will differ from $\widetilde{\overline{\theta}} = E_\theta(\theta|Y)$. We demonstrate in section 2 that, when $n_{irf} > n_p$, Θ is of measure zero in $\Re^{n_{irf}}$ and hence, with probability one, there is no $\widetilde{\overline{\lambda}} \in \Lambda$ such that $\widetilde{\overline{\theta}} = g(\widetilde{\overline{\lambda}})$. Thus, the two estimators differ in general. Only $\widehat{\overline{\theta}}$ is the Bayes estimator in this setting. Analogous arguments apply to the median response function under additively separable absolute loss.

In section 3, we discuss joint estimation and inference under Dirac delta loss. This approach was originally proposed by Inoue and Kilian (2013, 2019). That study showed how to rank posterior draws for structural VAR models based on the joint posterior density of the structural impulse responses when $n_{irf} = n_p$. Closed-form solutions for this density can be derived analytically, facilitating the implementation of this approach. Inoue and Kilian (2013) did not consider the case of $n_{irf} < n_p$ or of $n_{irf} > n_p$. In this paper, we generalize this approach to allow for arbitrary combinations of n_{irf} and n_p. In addition, we allow for the possibility that some or all of the response functions may be cumulated. We also extend the original analysis to allow for nonrecursive models with short-run exclusion restrictions, long-run exclusion restrictions, or combinations of sign and exclusion restrictions, as discussed in Arias, Rubio-Ramirez and Waggoner (2018). Especially, the latter situation is increasingly important in applied work for modeling the transmission of global

shocks to domestic economies and for modeling economies with frictions (e.g., Mumtaz and Surico 2009; Mountford and Uhlig 2009; Aastveit, Bjørnland and Thorsrud 2015; Beaudry, Nam and Wang 2018; Kilian and Zhou 2020a,b). In contrast, the original analysis in Inoue and Kilian (2013) only allowed for recursively identified models and models identified by sign restrictions. Finally, our analysis also allows for narrative restrictions, as discussed in Antolin-Diaz and Rubio-Ramirez (2018). Such inequality restrictions are increasingly used in sign-identified VAR models (e.g., Kilian and Zhou 2020a,b). We achieve this added generality by exploiting a new approach to deriving a closed-form solution for the joint posterior density function of the structural impulse responses that builds on results in Arias, Rubio-Ramirez and Waggoner (2018).

In Section 4 we propose a new loss function, referred to as angular loss, that is robust to the scale of the data as well as the scale of the structural shocks used in defining the impulse responses. We derive the Bayes estimator and joint credible sets based on this loss function. In our empirical applications, the estimates implied by this loss function are quite similar to those for other loss functions. We also address the concern that the optimal estimator may not be unique, adapting a proposal by Chernozhukov, Hong and Tamer (2007). Empirical applications suggest that nonuniqueness is not a problem in practice.

Section 5 contains two empirical illustrations that illustrate the use of the methods proposed in this paper. Our empirical examples illustrate that the standard practice of reporting pointwise impulse response error bands based on vectors of the $\alpha/2$ and $1-\alpha/2$ probability quantiles of the marginal posterior distributions of the impulse responses tends to grossly understate the estimation uncertainty about θ. We also find that there may be differences, in practice, between posterior median (and mean) response functions and the Bayes estimators of θ under absolute and under quadratic loss. The concluding remarks are in section 6.

2 Joint inference under absolute and under quadratic loss

The Bayesian estimate of a parameter is defined as the parameter value that minimizes in expectation the loss of the user. In the interest of clarity, we note that the choice of the loss function for impulse response analysis is subjective and that estimators derived under one loss function need not be optimal under another. Our point in this paper is not to argue for one loss function over another. Instead, in this section and the next, we present results for three commonly used loss functions (absolute, quadratic and Dirac delta loss). We leave it to the applied user to choose among these loss functions.

It is useful to first discuss joint impulse response inference under absolute and under quadratic loss, before turning to the Dirac delta loss function. The techniques discussed in this section may be applied to impulse responses in scalar autoregressions as well as in structural vector autoregressions. Let θ denote the n_{irf}-dimensional vector of unknown impulse responses obtained by appropriately stacking the impulse responses of interest. In a fully identified n-dimensional structural VAR models, for example, $n_{irf} = n^2(H+1)$, since there are n^2 impulse response functions for horizon $h = 0, 1, 2, ..., H$. $\theta \in \Theta$, where Θ is the set of all structural impulse response vectors that satisfy the identifying restrictions imposed on the structural VAR model.

We abstract from the details of the construction of θ, which may differ from one structural VAR model to another, because our analysis does not depend on these details (see Kilian and Lütkepohl 2017). The space Θ may be approximated by drawing from the posterior of the model parameters and simulating the distribution of the structural impulse responses that are consistent with the identifying assumptions. We first derive Bayes estimators of this impulse response vector under absolute and under quadratic loss. We then show how to construct joint credible sets for this estimator under the same loss function.

2.1 Constructing the impulse response estimator

Let $L(\theta, \overline{\theta})$ denote a loss function that maps $\Theta \times \Theta$ to $\Re$ such that

$$\widehat{\overline{\theta}} \;=\; \text{argmin}_{\overline{\theta} \in \Theta} E_{\theta}[L(\theta, \overline{\theta})], \tag{1}$$

where $\overline{\theta} \in \Theta$ denotes an action. It is assumed that there exists a unique minimum and that the expectation is with respect to the joint posterior distribution of θ over Θ. Our analysis in this section focuses on the loss functions

$$L(\theta, \overline{\theta}) \;=\; \begin{cases} \sum_{j=1}^{n_{irf}} (\theta_j - \overline{\theta}_j)^2 & \text{(quadratic loss)} \\ \sum_{j=1}^{n_{irf}} |\theta_j - \overline{\theta}_j| & \text{(absolute loss)} \end{cases}, \tag{2}$$

where θ_j and $\overline{\theta}_j$ are the jth elements of θ and $\overline{\theta}$, respectively.

In practice there is no closed-form solution to (1), so we need to approximate the solution by simulation. Then the Bayes estimator is given by

$$\widehat{\overline{\theta}}_M \;=\; \text{argmin}_{\overline{\theta} \in \widehat{\Theta}_M} \frac{1}{M} \sum_{i=1}^{M} L(\theta^{(i)}, \overline{\theta}), \tag{3}$$

where $\theta^{(i)}$ is the ith posterior draw and $\widehat{\Theta}_M$ consists of M posterior draws. Specifically, under the quadratic (QL) and under the absolute loss function (AL), respectively,

$$\widehat{\overline{\theta}}_M^{QL} \;=\; \text{argmin}_{\overline{\theta} \in \widehat{\Theta}_M} \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{n_{irf}} (\theta_j^{(i)} - \overline{\theta}_j)^2, \tag{4}$$

$$\widehat{\overline{\theta}}_M^{AL} \;=\; \text{argmin}_{\overline{\theta} \in \widehat{\Theta}_M} \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{n_{irf}} |\theta_j^{(i)} - \overline{\theta}_j|, \tag{5}$$

where $\theta_j^{(i)}$ denotes the jth element of the ith draw of θ, denoted $\theta^{(i)}$, from its posterior distribution.

2.2 Comparison with conventional posterior median (or mean) impulse response functions

The key difference from the conventional median or mean response functions in the literature is that the solution $\widehat{\overline{\theta}}_M$ is always contained in the set of structural impulse responses that can be obtained by drawing from the posterior of θ. This is true for any combination of p and H. In contrast, in the empirically relevant case when the number of estimated structural responses (n_{irf}) exceeds the number of estimated structural model parameters (n_p), the posterior median response function will not necessarily be contained in Θ, even as the number of posterior draws approaches infinity. The reason is that in this case the posterior distribution of the structural impulse response functions is defined over a lower-dimensional manifold. Our analysis recognizes that the space of θ may be a lower dimensional smooth manifold, whereas the conventional approach postulates that it is the Euclidian space. This fact does not change the posterior distribution for the structural model parameters, but it affects inference about θ.

Even when $n_{irf} \leq n_p$, the vector of marginal posterior medians will coincide with $\widehat{\overline{\theta}}_M$, only as the number of posterior draws approaches infinity. When $n_{irf} > n_p$, the vector of pointwise medians obtained from the marginal posterior distributions does not solve expression (1) even in the limit. Thus, the posterior median response function is not the Bayes estimator of θ under absolute loss. Analogous results hold under quadratic loss for the mean response function, defined as the vector of pointwise means from the marginal posterior impulse response distributions. This point is illustrated by the following stylized example.

Consider estimating the scalar-valued Gaussian AR(1) model

$$y_t = \rho y_{t-1} + u_t, \tag{6}$$

$$u_t \overset{iid}{\sim} N(0,1), \tag{7}$$

from the observations $y_0, y_1, ..., y_T$ with a diffuse conjugate Gaussian prior. Then, conditional on y_0, the posterior distribution of ρ is

$$\rho \sim N\left(\rho_{mean}, \frac{1}{n_T}\right), \tag{8}$$

where $\rho_{mean} = \sum_{t=1}^{T} y_t y_{t-1} \Big/ \sum_{t=1}^{T} y_t^2$ and $n_T = \sum_{t=1}^{T} y_{t-1}^2$.

2.2.1 The Conventional Approach

For expository purposes, we consider the reduced-form impulse response function. With a slight abuse of notation, denote this impulse response vector by θ. Focusing on the reduced-form responses allows us to isolate the essence of the problem with the conventional approach, which are unrelated to the uncertainty about the error variance. Since the response to a unit shock at horizon 0 is known to be 1, we focus on higher horizons. Suppose that we are interested in joint inference at horizons 1 and 2 only. Then $\theta = [\rho,\ \rho^2]$. In the context of this example, the condition that $n_{irf} > n_p$ in a structural VAR model translates to the number of estimated reduced-form responses, which is 2 in this example, exceeding the lag order of 1.

We consider first the posterior mean response function and then the posterior median response function. The posterior mean response function is

$$\left[\rho_{mean},\ \rho_{mean}^2 + \frac{1}{n_T}\right]. \tag{9}$$

By construction, there is no possible value of ρ in the AR(1) model such that $\theta = [\rho,\ \rho^2]$ produces the posterior mean response function because $1/n_T > 0$.

Next, let ρ_{median} denote the posterior median of ρ, where $\rho_{median} = \rho_{mean}$ under our

assumptions. Because

$$\begin{aligned} P(\rho^2 \leq x) &= P(-\sqrt{x} \leq \rho \leq \sqrt{x}) \\ &= \Phi\left(\sqrt{n_T}(\sqrt{x} - \rho_{median})\right) - \Phi(\left(\sqrt{n_T}(-\sqrt{x} - \rho_{median})\right), \end{aligned} \tag{10}$$

it is always true that

$$P(\rho^2 \leq \rho^2_{median}) \ < \ \frac{1}{2}. \tag{11}$$

In other words, the square of ρ_{median} is not the posterior median of ρ^2. Thus, the posterior median response function can be written as

$$\left[\rho_{median},\ \rho^2_{median} + \eta\right], \tag{12}$$

where $\eta > 0$. By inspection, there is no value of ρ in the AR(1) model such that the posterior median response function matches the actual impulse response function $\theta = [\rho,\ \rho^2]$. In contrast, the Bayes estimator $\widehat{\overline{\theta}}_M$ satisfies this constraint by construction. Note that these results do not depend on the choice of the prior or the specification of the Bayesian model.

The fact that the conventional posterior mean and median response functions in this example fail to recover $\widehat{\overline{\theta}}_M$ calls into question the claim that the posterior median response function is necessarily optimal under the loss function $\sum_{i=1}^{k} |\theta_i - \overline{\theta}_i|$ (see Baumeister and Hamilton 2018). The reason that we reach a different conclusion is not that we changed the loss function. Rather the difference is about the range over which to integrate when computing the expected loss. Whereas we restrict ourselves to the set of impulse response functions contained in Θ, the conventional approach considers in addition infeasible solutions not contained in Θ.[5]

[5] It may be tempting to argue that the conventional solution for θ is preferred because in this AR(1) example it generates the same expected loss at $h = 1$ and lower expected loss at $h = 2$. This argument would be misleading because the conventional solution only reduces the expected loss by violating the constraint that solutions for θ must

2.2.2 Why mean and median response functions may be misleading

It may seem that it is a matter of taste whether to restrict the minimization problem for θ to elements of Θ or not. We would argue that it is not. The reason why we do not find the conventional formulation of the minimization problem underlying the Bayes estimator of θ compelling is that the construction of the statistics reported in econometrics should reflect the preferences of applied users. For example, it is widely recognized there is no point in econometricians exploring loss functions that are at odds with the objectives of applied users. Likewise, there is no point in defining the space of possible solutions in impulse response analysis in a way that conflicts with the objectives of applied users.

In many studies it is the shape of the response function that users of structural VAR models are most interested in. For example, macroeconomists are often interested in whether the response function for real output to demand shocks is hump shaped or not (e.g., Cochrane 1994). Likewise, many macroeconomists have abandoned frictionless neoclassical models and adopted models with nominal or real rigidities based on VAR evidence of sluggish or delayed responses of inflation and output (e.g., Woodford 2003). Similarly, users of structural VAR models in international economics tend to be interested in whether there is delayed overshooting in the response of the exchange rate to monetary policy shocks (e.g., Eichenbaum and Evans 1995).

Moreover, many empirical questions can only be answered by considering several impulse response functions in conjunction. For example, the question of whether a global oil price shock creates stagflation in the domestic economy by necessity involves studying the responses of inflation as well as real output. Likewise, researchers interested in the effects of a U.S. monetary policy shock care about the responses of both real output and inflation because the loss function of the Federal Reserve depends on both real output and inflation, making it necessary to consider these

be feasible.

response functions jointly. Another good example is Blanchard (1989). This study uses a macroeconomic VAR model to evaluate the implication of standard Keynesian models that (1) positive demand innovations increase output and decrease unemployment persistently, and (2) that a favorable supply shock triggers an increase in unemployment without a decrease in output. Similarly, Sims and Zha (1999) stress that it is often thought that plausible patterns of the response of the economy to a disturbance in monetary policy should have interest rates rising, the money stock falling, output falling and prices falling.

Thus, it is fair to conclude that most applied researchers are interested in the shape of impulse response functions and in the comovement across impulse response functions. This fact is important because the rationale for median and mean response functions is based on the implicit premise that applied users are not interested in the shape of a given impulse response function (or in the comovement across impulse response functions). It is this premise that motivates the conventional approach of minimizing the expected loss of the response function θ over the Euclidian space rather than over Θ.

Our concern with this approach is that it is not possible to answer questions about the shape of a given impulse response function based on statistics that do not preserve the dynamics of the impulse response function implied by the underlying autoregressive model. Likewise, it is not possible to evaluate the comovement across the response functions contained in θ based on statistics that do not replicate the comovement implied by the underlying autoregressive model. As a result, computing median or mean response functions may mislead the applied user about the shape of the response function, which is of foremost interest in applied work, and more generally the comovement across response functions in multivariate models. This point is best illustrated by example.

Consider the same AR(1) model that served as our example earlier. As before, the posterior distribution of the slope parameter ρ is $N\left(\rho_{mean}, \frac{1}{n_T}\right)$, conditional on the data The impulse re-

sponse horizons of interest are $h = 1, 2, 3, 4$. Then the posterior mean of $\theta = \left[\rho, \rho^2, \rho^3, \rho^4\right]$ is given by

$$\left[\rho_{mean}, \rho_{mean}^2 + 1/n_T, \rho_{mean}^3 + 3\rho_{mean}/n_T, \rho_{mean}^4 + 6\rho_{mean}^2/n_T + 3/n_T^2\right],$$

where we make use of the fact that $\rho = \rho_{mean} + z/\sqrt{n_T}$ with $z \sim N(0, 1)$. For example, let $\rho_{mean} = 0.7$ and $n_T = 5$. Then the mean impulse response function is $[0.70, 0.69, 0.76, 0.95]$. In other words, the posterior mean response function implies that the response is initially falling, but increasing at longer horizons. This pattern is inconsistent with any value $0 < \rho < 1$ including $\rho = \rho_{mean} = 0.7$ because in this case the response function is strictly decreasing with the horizon. It is also inconsistent with $\rho = 1$ or $\rho = 0$ because in this case the response function is flat and with $\rho > 1$ because in this case the response function is strictly increasing with the horizon. Finally, it is inconsistent with $\rho < 0$ because the mean response is positive at horizon 1. Thus, there is no possible value of ρ that could generate the posterior mean response function. This makes the mean response function misleading as an estimator of the vector θ.

Similar problems arise with the posterior median response function. Figure 1 shows an example, in which the median response function follows an irregular pattern that is clearly inconsistent with the shape of the impulse response function in the AR(1) model for any possible value of ρ. A user of this posterior median response function would be tempted to conclude that the response function has a shape that is clearly inconsistent with the posited AR(1) model.

These two examples show that mean and median response functions cannot be trusted to capture the shape of the impulse response function. Since in higher-dimensional models all impulse response functions of interest are stacked into the vector θ, the same concern applies in the structural VAR setting. Clearly, one cannot trust median or mean response function to capture the dynamics of θ in structural VAR(p) models, when these statistics cannot even capture the shape of the response

function in the much simpler AR(1) model.[6]

2.3 Extensions to partially identified models

Although so far we have focused on fully identified structural VAR models, our analysis can be easily adapted to structural VAR models, in which only a subset of the structural shocks is identified. The only change is the dimensionality of the θ vector when evaluating the expected loss.[7] Otherwise, the implementation of our methods is unchanged. Like in the fully identified structural VAR model, the shape of the median response function may be inconsistent with the shape of the impulse response function implied by any possible parameterization of the underlying autoregressive model, when there are more structural impulse responses to be estimated than structural model parameters. The latter situation arises commonly in applied work. For example, Eichenbaum and Evans (1995) specify a recursive semi-structural model of monetary policy shocks with $n = 5$ variables, $p = 6$ autoregressive lags, and $H = 35$.

2.4 Joint credible sets

There has been growing interest in recent years in quantifying the joint uncertainty of vectors of VAR impulse response estimates. There is no consensus in the Bayesian literature on how to construct such credible sets. For example, Berger (1985, p. 145) does not view credible sets as having a clear decision-theoretic role at all, and therefore is leery of 'optimality' approaches to the selection of a credible set. He views credible sets as an easily reportable crude summary of the

[6] It should be noted that it is possible to further weight individual elements of θ in the loss function, depending on the variable of interest and the horizon of the response function. For example, a user may want to assign more weight to impulse responses at horizons of interest to policymakers. We do not consider this possibility here, since this option has not been considered in applied work. We do note, however, that the conventional median response function solves the loss minimization problem under additively separable absolute loss regardless of the weights applied to the elements of θ, whereas our approach will generate different solutions under absolute loss as well as quadratic loss, as the weights are varied.

[7] For example, when only one structural shock is identified, as is often the case in structural VAR models of monetary policy shocks, with this shock ordered last, θ is of length $n(H+1)$.

posterior distribution.[8]

In this section, we show how to construct joint credible sets for $\widehat{\overline{\theta}}_M$ under quadratic and under absolute loss as lowest posterior risk regions. This approach was previously discussed in Bernardo (2010). Our contribution is to make Bernardo's approach operational in the context of structural impulse response analysis under a range of alternative assumptions about the loss function. We define the $(1-\alpha)100\%$ joint credible set based on a loss function $L(\theta, \overline{\theta})$ by

$$\Theta_{1-\alpha,L} \;=\; \{\overline{\theta} \in \Theta : E_\theta(L(\theta, \overline{\theta})) \leq c_{1-\alpha,L}\}, \tag{13}$$

where $c_{1-\alpha,L}$ is the smallest number such that the posterior probability of $\Theta_{1-\alpha,L}$ is $1-\alpha$ and L refers to the loss function. This credible set can be estimated as

$$\widehat{\Theta}_{1-\alpha,M,QL} \;=\; \left\{ \overline{\theta} \in \widehat{\Theta}_M : \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n_{irf}} (\theta_j^{(i)} - \overline{\theta}_j)^2 \leq c_{1-\alpha,QL} \right\}, \tag{14}$$

$$\widehat{\Theta}_{1-\alpha,M,AL} \;=\; \left\{ \overline{\theta} \in \widehat{\Theta}_M : \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n_{irf}} |\theta_j^{(i)} - \overline{\theta}_j| \leq c_{1-\alpha,AL} \right\}, \tag{15}$$

where $\widehat{\overline{\theta}}_M^{QL}$ and $\widehat{\overline{\theta}}_M^{AL}$ denote the respective solutions to (3) for the loss functions in (2), and where $c_{1-\alpha,QL}$ and $c_{1-\alpha,AL}$ are the smallest values such that $\widehat{\Theta}_{1-\alpha,M,QL}$ and $\widehat{\Theta}_{1-\alpha,M,AL}$, respectively, have posterior probability $1-\alpha$ in the limit, as the number of posterior draws approaches infinity. Unlike the conventional credible sets used in the literature, our joint credible sets are designed to contain only members of the set of structural response functions. In practice, these joint credible set may be constructed by sorting $\frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)}, \theta^{(1)})$, $\frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)}, \theta^{(2)})$,... $\frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)}, \theta^{(M)})$, in ascending order and retaining the first $(1-\alpha)100\%$ draws, starting with the draw with the lowest

[8] For further discussion see Berger (1985, subsection 4.3.3) and Bernardo and Smith (1994, subsection 5.1.5), for example.

value.

2.5 Comparison with conventional error bands

The conventional approach in the literature has been to report posterior median (or mean) response functions along with pointwise error bands constructed from the upper and lower quantiles of the marginal posterior distributions of the structural impulse responses. While this approach makes sense when conducting inference on individual impulse responses, it is readily apparent that this approach is misguided if we are interested in inference on the vector θ, regardless of whether n_{irf} exceeds n_p or not.

Our premise in this paper is that applied researchers are interested in the shape of individual response function and in the comovement across response functions, which necessarily involves conducting inference simultaneously across multiple horizons and potentially for multiple variables. Constructing error bands by stringing together quantiles of the marginal posterior distributions of individual impulse responses, is akin to arguing that in studying a system of regression equations it is sufficient to base inference on t-test statistics for each coefficient in a given equation, ignoring that the t-statistics are dependent within and across equations. It is immediately obvious that this approach fails to capture the uncertainty about the vector θ, given that it ignores the dependence across the elements of θ. The only defense of this approach is that there has been no constructive alternative to the use of such pointwise error bands until recently. As we have shown, there are easy-to-implement alternative methods that avoid this drawback, rendering the use of pointwise error bands obsolete.

By restricting all members of the joint credible set for θ to lie in the set Θ, we not only ensure that each member of that set is feasible, but we also make sure that $\overline{\theta}$ corresponds to a draw from the joint distribution of the impulse responses that satisfies the dependence structure of the impulse

responses. Even when posterior median (or mean) response functions are numerically close to $\widehat{\overline{\theta}}_M$, these joint credible sets will not resemble conventional pointwise error bands. Rather, they will look like shot-gun trajectories employed in ballistics, with each structural model represented by a set of n^2 such trajectories up to horizon H.[9] This feature addresses the concern raised by Sims and Zha (1999) that we need to be able to visualize likely departures from the baseline estimate of θ. Similar techniques have been employed in Inoue and Kilian (2013, 2016) and a number of applied studies to characterize the joint estimation uncertainty in structural VAR models.

Of course, in practice, M is typically so large that we cannot distinguish individual response functions or, for that matter, responses from different structural models. The shotgun trajectory plot does, however, contain the information required to make these assessments. If we are interested in evidence that a positive oil price shocks causes stagflation in the U.S. economy, for example, we may display the response functions that satisfy the definition of a stagflationary response in a different color than response functions that do not, as illustrated in Inoue and Kilian (2016). Similarly, if we are interested in whether the most likely models satisfy the requirement that a monetary policy shock raises the interest rate, lowers the price level and lowers real output, as discussed by Sims and Zha (1999), we can plot the responses for this subset of structural models in a different color. We can also make probability statements about the conditional comovement of several response functions contained in the joint credible set. Finally, if we are interested in the shape of a given impulse response function, we can directly assess whether the response function is hump-shaped or whether there is delayed overshooting, for example, by highlighting these responses in a different color. We illustrate how information about the shape and comovement of the response functions can be visualized in section 5. Further discussion of techniques for extracting such information

[9] These trajectories are also conceptally similar to the forecast paths reported in weather forecasting, when outlining the predicted path of a hurricane.

from shot-gun trajectory plots can be found in Inoue and Kilian (2016).[10]

It may be tempting to construct two-dimensional impulse response error bands from these joint credible sets. For example, one might wish to bound the joint credible set by constructing the upper and lower envelope around the response estimates in the credible region.[11] Such error bands may indeed help summarize the range of uncertainty about the impulse responses, but it is important to understand that they would not be a good substitute for the plot of the trajectories. First, these error bands, much like the conventional pointwise error bands, need not be contained in the set Θ of feasible solutions. Second, by focusing on error bands rather than the shot-gun trajectory plots, we suppress the information that applied users are most interested in. Consider the example of a structural VAR model of monetary policy shocks. For starters, we lose the information required to judge whether the response of output to a monetary policy shock is hump-shaped. Next, as stressed by Sims and Zha (1999), when looking at the response of output to a monetary policy shock, we need to know whether a stronger output decline within the credible set remains consistent with the implications of a monetary policy shock for the other model variables. This question cannot be answered based on error bands. It can only be answered based on the information contained in the shot-gun trajectory plots.

[10] One may object that if we care about features such as hump-shaped responses, we could instead use a loss function that reflects this objective. This would allow the user to directly evaluate the posterior support for a hump shape. In practice, however, one cannot target all possible objectives a potential user may have in mind. It therefore makes sense for researchers to settle on default conventions for reporting impulse response estimates that allow users with different objectives to extract as much information as possible from the impulse response plot. This is indeed the approach typically taken both in the Bayesian and in the frequentist structural VAR literature (e.g., Montiel Olea and Plagborg-Møller 2019). Moreover, it is often the impulse response plots that first suggest what feature might be interesting to consider.

[11] It should be noted that when $n_{irf} > n_p$, one cannot rule out that the credible region is disjoint Identifying disjoint regions is a challenge in practice for computational reasons. One never knows whether a region is truly disjoint or whether the number of posterior draws is too small. This is especially true near the edges of the credible region. If one were to plot the envelope ignoring any disjoint regions within this band, the resulting joint error band would still have the intended probability content, but would effectively be greater than desirable when there are disjoint regions. This problem is not unique to this approach. Similar problems afflict other two-dimensional error bands.

3 Joint inference under Dirac delta loss

Although the traditional approach in applied studies has been to focus on absolute or quadratic loss functions, respectively, starting with Inoue and Kilian (2013, 2019) there has been growing interest in the application of the Dirac delta loss function to structural impulse response analysis.[12] Unlike the quadratic and absolute loss functions considered in section 2, the Dirac delta loss function is not additively separable. We assume for now that the structural VAR model is fully identified. Let θ denote the $n^2(H+1)\times 1$ vector of structural impulse response functions that satisfy the identifying restrictions, where $\theta \in \Theta$. As before, Θ is a lower-dimensional smooth manifold when the number of structural impulse responses to be estimated exceeds the number of structural model parameters used to construct these responses. We are interested in minimizing in expectation the Dirac delta loss function, $L(\theta,\overline{\theta}) = -\delta(\overline{\theta}-\theta)$, where $\delta(\cdot)$ denotes the Dirac delta function. This problem is equivalent to maximizing the joint posterior density of θ, which occurs when $\overline{\theta}$ is the posterior mode of the joint posterior density.[13]

3.1 The most likely impulse response estimate and joint HPD credible sets

The procedure proposed in Inoue and Kilian (2013, 2019) for finding the modal posterior draw is simple. We first rank the M posterior draws for θ that satisfy the identifying restrictions of the structural VAR model based on the value of the joint posterior density function $f(\theta|y_1,...,y_T)$. Then the modal (or most likely) posterior draw for θ, denoted $\widehat{\overline{\theta}}_M^{DDL}$, where DDL stands for Dirac delta loss, will be the draw with the highest joint posterior value. By construction, the mode of the joint posterior distribution will be within Θ.

[12] This approach has been applied in a number of recent studies including Herwartz and Plödt (2016), Herrera and Rangaraju (2020), Zhou (2020), Cross, Nguyen and Tran (2019) and Rausser and Stürmer (2020).

[13] The use of the mode has a long tradition in Bayesian inference, as does the use of highest posterior density (HPD) regions for impulse response inference (e.g., Koop 1996; Zha 1999; Waggoner and Zha 2012; Plagborg-Møller 2019). Similarly, Sims and Zha (1999) suggest that impulse response inference should focus on deviations from the baseline estimate that are "most likely".

The corresponding $1-\alpha$ probability joint credible sets may be constructed by retaining the $(1-\alpha)100\%$ draws for θ with the highest joint posterior density values among the M draws that satisfy the identifying restrictions. The most likely impulse response estimate thus may be viewed as the limit of a joint $1-\alpha$ probability highest-posterior density (HPD) credible set. As in section 2, the plot of this joint credible set resembles a shot-gun trajectory plot and may be evaluated in the same way.

The main challenge, in practice, is how to compute the values of the joint posterior density of the structural impulse responses for each posterior draw. In our earlier work, we analytically derived a closed-form solution for the function $f(\cdot)$ that makes this approach accurate and computationally efficient. In this section, we propose an alternative derivation of this density that is substantially more general than the expressions derived in Inoue and Kilian (2013, 2019) under the restrictive assumption that $n_{irf} = n_p$.

3.2 The joint posterior density of the structural impulse responses

Following Arias et al. (2018), consider a pth-order structural VAR model

$$\begin{aligned} y_t' A_0 &= \sum_{j=1}^{p} y_{t-j}' A_j + c + \varepsilon_t' \\ &= x_t' A_+ + \varepsilon_t', \quad \text{for } t = 1, 2, ..., T, \end{aligned} \tag{16}$$

where y_t is an $n \times 1$ vector of observed variables, A_j is an $n \times n$ matrix of parameters for $j = 0, 1, ..., p$, c is a $1 \times n$ vector of intercepts, ε_t is an $n \times 1$ vector of structural shocks, $x_t' = [y_{t-1}' \; y_{t-2}' \; \cdots \; y_{t-p}' \; 1]$ and $A_+' = [A_1' \; \cdots \; A_p' \; c']$. A_0 is assumed to be nonsingular. The implied reduced-form VAR representation is

$$y_t' = x_t'B + u_t', \quad \text{for } t = 1, 2, ..., T, \tag{17}$$

where $B = A_+A_0^{-1}$, $u_t' = \varepsilon_t'A_0^{-1}$ and $E(u_tu_t') = \Sigma = (A_0A_0')^{-1}$. The structural VAR model can also be written in the orthogonal reduced-form parameterization characterized by the reduced-form parameters B and Σ and an orthogonal matrix Q:

$$y_t' = x_t'B + \varepsilon_t'Q'P, \quad \text{for } t = 1, 2, ..., T, \tag{18}$$

where P is the lower triangular Cholesky decomposition of Σ with the diagonal elements normalized to be positive such that $PP' = \Sigma$ and $QQ' = I_n$. Q is also referred to as the rotation matrix. There is a one-to-one mapping between the parameters (B, Σ, Q) and (A_0, A_+).

The further analysis differs depending on the nature of the identifying restrictions. We first discuss the case of structural VAR models that are set-identified by sign restrictions. We then consider the case of structural VAR models exactly identified by any combination of short-run and long-run exclusion restrictions.

3.2.1 Models identified by sign restrictions

The standard approach in the literature is to define a normal-inverse Wishart (NIW) distribution over the reduced-form parameters characterized by four parameters: a scalar $\nu \geq n$, an $n \times n$ symmetric and positive definite matrix Φ, an $n(p+1) \times n$ matrix Ψ, and an $n(p+1) \times n(p+1)$ symmetric and positive definite matrix Ω. The density

$$NIW_{(\nu,\Phi,\Psi,\Omega)}(B, \Sigma) \propto |\det(\Sigma)|^{\frac{\nu+n+1}{2}} e^{-\frac{1}{2}tr(\Phi\Sigma)^{-1}} |\det(\Sigma)|^{-\frac{m}{2}} e^{-\frac{1}{2}vec(B-\Psi)'(\Sigma\otimes\Omega)^{-1}vec(B-\Phi)} \tag{19}$$

can be expressed up to scale as the product of an inverse Wishart density for Σ and a conditionally normal density for B (see Arias et al. 2018). Conditional on the reduced-form parameters, the density over the set of orthogonal matrices Q is uniform. This prior is conjugate in that the corresponding posterior density is of the same functional form.

The key technical difference from Inoue and Kilian (2013, 2019) is that in the current paper we start with the posterior of A_0 and A_+, as defined in Arias et al. (2018), whereas in our earlier work we applied the change-of-variable method to the joint posterior of the reduced-form slope parameters, the error covariance matrix and the rotation matrix. Given a uniform-normal-inverse-Wishart family of prior densities for (B, Σ, Q) with parameters $\bar{\nu}$, $\bar{\Phi}$, $\bar{\Psi}$ and $\bar{\Omega}$, the posterior density of A_0 and A_+ is given by

$$NGN_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})} \propto |\det(A_0)|^{\tilde{\nu}-n} e^{-\frac{1}{2}\text{vec}(A_0)'(I_n\otimes\Phi)\text{vec}(A_0)} \times e^{-\frac{1}{2}\text{vec}(A_+-\tilde{\Psi}A_0)'(I_n\otimes\Omega)^{-1}\text{vec}(A_+-\tilde{\Psi}A_0)}, \tag{20}$$

where $\tilde{\nu} = T + \bar{\nu}$, $\tilde{\Omega} = (X'X + \bar{\Omega}^{-1})^{-1}$, $\tilde{\Psi} = \tilde{\Omega}(X'Y + \bar{\Omega}^{-1}\bar{\Psi})$, $\tilde{\Phi} = Y'Y + \bar{\Phi} + \bar{\Psi}'\bar{\Omega}^{-1}\bar{\Psi} - \tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\Psi}$, $Y = [y_1 \ \cdots \ y_T]'$ and $X = [x_1 \ \cdots \ x_T]'$ (see equation 2.8 in Arias et al. 2018).[14]

Let Θ_h denote the structural impulse response matrix at horizon h implied by (16). Let J_{NGN} denote the $(H+1)n^2 \times n^2(\min(H,p)+1)$ Jacobian matrix of the possibly cumulated impulse responses in $\theta = \text{vec}([\Theta_0 \ \Theta_1 \ \cdots \ \Theta_H])$ with respect to $\text{vec}([A_0 \ \ A_1 \ \cdots \ A_p])$. The following proposition, which is a corollary of the results in Arias et al. (2018), states the joint posterior

[14] The approach of specifying a marginally uniform prior on the rotation matrix has been standard in the literature on sign-identified structural VAR models (see, e.g., Rubio-Ramirez et al. 2010; Arias et al. 2018; Antolin-Diaz and Rubio-Ramirez 2018). As is well known, this prior may be inadvertently informative about the structural impulse responses. The practical importance of this problem is examined in Inoue and Kilian (2020). An alternative approach that avoids specifying a prior for the orthogonal matrix Q is discussed in Giacomini and Kitagawa (2020). Another way to avoid the specification of the Q prior is to specify a prior directly on the parameters of the structural moving average representation, as discussed in Plagborg-Møller (2019).

density of θ.

Proposition 1:

$$f(\theta|y_1,...,y_T) \;=\; NGN_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}|J'_{NGN}J_{NGN}|^{\frac{1}{2}}. \tag{21}$$

This result follows from Arias et al.'s (2018) Theorem 2 and their equation 2.8. If the number of estimated impulse responses (n_{irf}) is no larger than the number of estimated structural parameters used in computing these structural responses, $f(\theta|y_1,...,y_T)$ simplifies to $NGN_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}|J_{NGN}|$.

To make our procedure operational, what remains to be derived is the Jacobian J_{NGN}, which can be written as the product of four Jacobian matrices. The first Jacobian allows for the fact that some impulse responses may be cumulated. Let S be the $n \times n$ diagonal selection matrix whose (i,i)th element is one if the impulse response functions of the ith variable are cumulative and is zero otherwise. Let Ξ denote the uncumulated impulse responses. Then the Jacobian of the cumulative impulse response functions $\text{vec}([\Xi_0 \; \Xi_0 + \Xi_1 \; \cdots \; \sum_{j=0}^{H} \Xi_j])$ with respect to noncumulative impulse response functions $\text{vec}([\Xi_0 \; \Xi_1 \; \cdots \; \Xi_H])$ is the $n^2(H+1) \times n^2(H+1)$ matrix

$$J_1 \;=\; \begin{bmatrix} I_{n^2} & 0_{n\times n} & \cdots & 0_{n\times n} \\ I_n \otimes S & I_{n^2} & \cdots & 0_{n\times n} \\ \vdots & \vdots & \ddots & \vdots \\ I_n \otimes S & I_n \otimes S & \cdots & I_{n^2} \end{bmatrix}. \tag{22}$$

If there are no responses to be cumulated, J_1 reduces to the identity matrix and may be dropped.

The second Jacobian matrix is that of $\text{vec}([\Xi_0 \; \cdots \; \Xi_H])$ with respect to $\text{vec}([A_0'^{-1} \; \Phi_1 \; \cdots \; \Phi_H])$, where Φ_h denotes the reduced-form impulse response matrix at horizon h. Because $\Xi_h = \Phi_h A_0'^{-1}$,

this Jacobian is given by the $n^2(H+1) \times n^2(H+1)$ matrix

$$J_2 \;=\; \begin{bmatrix} I_{n^2} & 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} \\ (I_n \otimes \Phi_1) & A_0^{-1} \otimes I_n & \cdots & 0_{n^2\times n^2} \\ \vdots & \vdots & \ddots & \vdots \\ (I_n \otimes \Phi_H) & 0_{n^2\times n^2} & \cdots & A_0^{-1} \otimes I_n \end{bmatrix}. \tag{23}$$

The third Jacobian matrix is that of $\text{vec}([A_0'^{-1}\ \Phi_1\ \cdots\ \Phi_H])$ with respect to $\text{vec}([A_0\ B_1\ \cdots\ B_{\min(p,H)}])$, where $B_i = A_i A_0^{-1}$, $i = 1, ..., p$. If $H \leq p$, this Jacobian is given by the $n^2(H+1) \times n^2(H+1)$ matrix

$$J_3 \;=\; \begin{bmatrix} -(A_0^{-1} \otimes A_0'^{-1})K_n & 0_{n^2\times n^2} & 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & I_{n^2} & 0_{n^2\times n^2} & \cdots & 0_{n^2\times n^2} \\ 0_{n^2\times n^2} & X(\Phi_2, B_1) & I_{n^2} & \cdots & 0_{n^2\times n^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{n^2\times n^2} & X(\Phi_H, B_1) & X(\Phi_H, B_2) & \cdots & I_{n^2} \end{bmatrix}. \tag{24}$$

where K_n is the $n \times n$ commutation matrix and $X(\Phi_i, B_j)$ denotes $n^2 \times n^2$ Jacobian matrix of $\text{vec}(\Phi_i)$ with respect to $\text{vec}(B_j)$. It follows from equation (A.24) in Cheng, Han and Inoue (2020) that

$$X(\Phi_i, B_j) \;=\; \sum_{k=0}^{i-1} (M'^{k'} \otimes \Phi_{i-1-k}) S, \tag{25}$$

where $M = [I_n\ 0_{n\times n(p-1)}]'$ and

$$F \;=\; \begin{bmatrix} B_1' & B_2' & \cdots & B_p' \\ 0_{n^2(p-1)\times n} & & I_{n(p-1)} & \end{bmatrix}, \tag{26}$$

such that $\text{vec}(B_j') = K_n \text{vec}(B_j)$, and $S_j = [0_{n^2(j-1)\times n^2}\ K_n\ 0_{n^2(p-j)\times n^2}]'$. If $H > p$, in contrast, the

third Jacobian is given by the $n^2(H+1) \times n^2(p+1)$ matrix

$$J_3 = \begin{bmatrix} -(A_0^{-1} \otimes A_0'^{-1})K_n & 0_{n^2 \times n^2} & 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & I_{n^2} & 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} \\ 0_{n^2 \times n^2} & X(\Phi_2, B_1) & I_{n^2} & \cdots & 0_{n^2 \times n^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{n^2 \times n^2} & X(\Phi_p, B_1) & X(\Phi_p, B_2) & \cdots & I_{n^2} \\ 0_{n^2 \times n^2} & X(\Phi_{p+1}, B_1) & X(\Phi_{p+1}, B_2) & \cdots & X(\Phi_{p+1}, B_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{n^2 \times n^2} & X(\Phi_H, B_1) & X(\Phi_H, B_2) & \cdots & X(\Phi_H, B_p) \end{bmatrix}. \quad (27)$$

The fourth Jacobian matrix is that of $\text{vec}([A_0\ B_1\ \cdots\ B_{\min(p,H)}])$ with respect to $\text{vec}([A_0\ A_1\ \cdots\ A_{\min(p,H)}])$ and is given by the $n^2(\min(p,H)+1) \times n^2(\min(p,H)+1)$ matrix

$$J_4 = \begin{bmatrix} I_{n^2} & 0_{n^2 \times n^2} & \cdots & 0_{n^2 \times n^2} \\ -(A_1 A_0^{-1} \otimes A_0'^{-1})K_n & (I_n \otimes A_0'^{-1})K_n & \cdots & 0_{n^2 \times n^2} \\ \vdots & \vdots & \ddots & \vdots \\ -(A_{\min(p,H)} A_0^{-1} \otimes A_0'^{-1})K_n & 0_{n^2 \times n^2} & \cdots & (I_n \otimes A_0'^{-1})K_n \end{bmatrix}. \quad (28)$$

Thus, the Jacobian $J_{NGN} = J_1 J_2 J_3 J_4$ is of dimension $n^2(H+1) \times n^2(\min(p,H)+1)$. This matrix becomes very large in typical macroeconomic applications. Although its determinant simplifies to $|A_0|^{-(H+1)n}$ when $H \leq p$, it does not simplify, when $H > p$. This poses a computational challenge. We work around this problem by using the LU decomposition and computing the sum of the log of the diagonal elements for the log of the Jacobian determinant.[15]

The implementation of this algorithm can be summarized in four steps:

[15] Our code utilizes the function logdet.m, provided on the Mathwork file exchange. © 2009, Dahua Lin.

1. Draw A_0 and A_+ from the NGN posterior distribution. From this draw, construct θ subject to the maintained sign restrictions.

2. Evaluate the NGN kernel in equation (21) at this posterior draw and multiply the kernel by $|J'_{NIW}J_{NIW}|^{1/2}$ where $|\cdot|$ denotes the determinant to obtain the log posterior value for this draw.

3. Repeat steps 1 and 2 M times and sort the posterior draws in descending order by the value of the log posterior computed in step 2.

4. The modal value of θ corresponds to the first element of the sorted log posterior values. The θ draws in the $100(1-\alpha)\%$ joint credible set are correspond to the first $[(1-\alpha)M]$ elements of the sorted log posterior values, where $[\cdot]$ denotes the integer part.

3.2.2 Models exactly identified by short-run and/or long-run exclusion restrictions

Consider the NIW family of priors for the reduced-form parameters. Suppose that there are $\frac{n(n-1)}{2}$ short-run and/or long-run exclusion restrictions of the form

$$C \begin{bmatrix} \text{vec}(A_0^{-1\prime}) \\ \text{vec}(B(1)^{-1}A_0^{-1\prime}) \end{bmatrix} = 0_{\frac{n(n-1)}{2}\times 1}, \tag{29}$$

where C is a $\frac{n(n-1)}{2} \times 2n^2$ matrix of constants and $B(1) = I_n - B_1 - \cdots - B_p$ is assumed to be invertible. The rank of C is assumed to be $n(n-1)/2$. Typically, each of the rows of C consists of a 1 and $2n^2 - 1$ 0's. We assume that there is a unique A_0 that satisfies $A_0^{-1\prime}A_0^{-1} = \Sigma$ given (29). In this case the joint posterior density of θ, defined to include the elements of θ that are restricted to be zero, can be derived directly from the posterior density of the reduced-form parameters, building on the results in Arias et al. (2018). The following proposition, which is not a special case

of Proposition 1, states the joint posterior density of θ (up to scale).

Proposition 2:

$$f(\theta|y_1,...,y_T) \;=\; NIW_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}|J_{NIW}'J_{NIW}|^{\frac{1}{2}}, \tag{30}$$

where $J_{NIW} = J_1J_2J_3J_4$, $J_4 = D^{-1}N$, D and N are the $n^2(\min(p,H)+1) \times n^2(\min(p,H)+1)$ and $n^2(\min(p,H)+1) \times \frac{n(n+1)}{2} + n^2\min(p,H)$ matrices such that

$$D \;=\; \begin{bmatrix} -D_n^+(A_0^{-1\prime} \otimes \Sigma + (\Sigma \otimes A_0^{-1\prime})K_n) & 0 \\ -C\begin{bmatrix} (A_0^{-1} \otimes A_0^{-1\prime})K_n \\ (A_0^{-1} \otimes B(1)^{-1}A_0^{-1\prime})K_n \end{bmatrix} & 0 \\ 0 & I \end{bmatrix}, \tag{31}$$

$$N \;=\; \begin{bmatrix} I & 0 \\ 0 & -C\begin{bmatrix} 0 \\ A_0^{-1}B(1)^{-1\prime} \otimes B(1)^{-1} \end{bmatrix} \\ 0 & I \end{bmatrix}, \tag{32}$$

$D_n^+ = (D_n'D_n)^{-1}D_n'$ and D_n is the $n^2 \times \frac{n(n+1)}{2}$ duplication matrix such that $\text{vec}(\Sigma) = D_n\text{vech}(\Sigma)$.

Proof: Because the posterior distribution of B and Σ is $NIW_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}$, we only need to find the Jacobian matrix J_{NIW} for transforming $[\text{vech}(\Sigma)' \; \text{vec}[B_1 \; \cdots \; B_p]']$ to $\theta = \text{vec}([\Theta_0 \; \Theta_1 \; \cdots \; \Theta_H])$, taking into account any zero restrictions in (29).

This Jacobian matrix is the product of four Jacobian matrices. The first three of these matrices are the same as J_1, J_2 and J_3 in Proposition 1, whereas J_4 is replaced by the Jacobian matrix of

vec($[A_0\ B_1\ \cdots\ B_p]$) with respect to $[vech(\Sigma)'\ \text{vec}([B_1\ \ldots\ B_p])']$. It follows from (29) that

$$C\begin{bmatrix} -(A_0^{-1}\otimes A_0^{-1\prime})K_n & 0 \\ -(A_0^{-1}\otimes B(1)^{-1}A_0^{-1\prime})K_n & A_0^{-1}\otimes B(1)^{-1\prime}\otimes B(1)^{-1} \end{bmatrix} = 0. \tag{33}$$

Thus the differential of vec($[A_0\ B_1\ \cdots\ B_p]$) can be written as

$$\begin{aligned} &-\begin{bmatrix} D_n^+(A_0^{-1\prime}\otimes\Sigma+(\Sigma\otimes A_0^{-1\prime})K_n) & 0 \\ -C\begin{bmatrix}(A_0^{-1}\otimes A_0^{-1\prime}))K_n\\ (A_0^{-1}\otimes B(1)^{-1}A_0^{-1\prime})K_n\end{bmatrix} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix}\text{vec}(A_0)\\ \text{vec}(B_1)\\ \vdots \\ \text{vec}(B_p)\end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ -C\begin{bmatrix}0\\ A_0^{-1}B(1)^{-1\prime}\otimes B(1)^{-1}\end{bmatrix} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix}\text{vech}(d\Sigma)\\ \text{vec}[dB_1\ \cdots\ dB_p]\end{bmatrix}. \end{aligned} \tag{34}$$

Thus, we have $J_4 = D^{-1}N$ and it follows from Theorem 2 of Arias et al. (2018) that the volume element is given by

$$|J_{NIW}'J_{NIW}|^{\frac{1}{2}} \tag{35}$$

where $J_{NIW} = J_1J_2J_3$.

This proposition extends the analysis in Inoue and Kilian (2013) to nonrecursive exactly identified models based on any combination of short-run and long-run exclusion restrictions, while relaxing the assumption that $n_{irf} = n_p$. It also covers, as a special case, models identified based on long-run exclusion restrictions only and models that are recursively identified by zero restrictions on the structural impact multiplier matrix. For $n_{irf} = n_p$, our results for the latter model match

the closed-form solution derived in Inoue and Kilian (2013, 2019).

The implementation of this algorithm can be summarized in four steps:

1. Draw B and Σ from their $NIW_{(\tilde{\nu},\tilde{\Phi},\tilde{\Psi},\tilde{\Omega})}$ posterior distribution. From this draw, construct θ subject to the maintained short-run and/or long-run exclusion restrictions.

2. Evaluate the NIW kernel in equation (19) at this posterior draw and multiply the kernel by $|J'_{NIW}J_{NIW}|^{1/2}$, where $|\cdot|$ denotes the determinant to obtain the log posterior value for this draw.

3. Repeat steps 1 and 2 M times and sort the posterior draws in descending order by the value of the log posterior computed in step 2.

4. The modal value of θ corresponds to the first element of the sorted log posterior values. The θ draws in the $100(1-\alpha)\%$ joint credible set are correspond to the first $[(1-\alpha)M]$ elements of the sorted log posterior values, where $[\cdot]$ denotes the integer part.

3.3 Discussion

Our analysis of the Dirac delta loss function contributes to the literature in three dimensions. First, whereas in our earlier work on fully identified recursive and sign-identified structural VAR models we focused on the case of $n_{irf} = n_p$, in the current paper, Propositions 1 and 2 allow for $n_{irf} \leq n_p$ as well as for $n_{irf} > n_p$. It should be noted that the first $n_{irf} < n_p$ responses of the modal model obtained from the joint posterior distribution derived under the assumption that there are n_p responses, will not in general coincide with the responses of the modal model derived under the assumption that $n_{irf} < n_p$ because the minimizer of the expected loss may change with the dimension of θ.

Second, Propositions 1 and 2 cover not only the recursively identified and sign-identified structural VAR models studied by Inoue and Kilian (2013). It also enables us to deal with structural VAR models for which the joint posterior density of the structural impulse responses has not been derived to date. One example is nonrecursive models based on short-run and/or long-run restrictions. Another example is models based on combinations of sign and zero restrictions in the structural impact multiplier matrix. Especially the latter extension is practically important. Applications of structural VAR models to the transmission of global shocks to the domestic economy tend to be block recursive (e.g., Cushman and Zha 1997; Zha 1999). When global and/or domestic shocks are identified by sign restrictions, this results in a structural VAR model that combines sign and zero restrictions on A_0^{-1} (e.g., Mumtaz and Surico 2009; Kilian and Zhou 2020a) Similar situations arise, more generally, in sign-identified models when the feedback from one variable to another is delayed by informational or institutional frictions (e.g., Mountford and Uhlig 2009; Aastveit, Bjørnland and Thorsrud 2015; Nam and Wang 2018; Kilian and Zhou 2020b).

When imposing additional zero restrictions on the structural impact multiplier matrix of the model, Proposition 1 may be applied with selected elements of A_0^{-1} restricted to be zero and included in θ. When evaluating a model that combines zero and sign restrictions on the structural impact multiplier matrix, the posterior draws must be reweighted based on the importance sampler described in Arias et al. This reweighting changes the joint posterior density. The new joint density may be constructed by reweighting the density in Proposition 1 by the importance sampling weights. This requires adding the log of the respective importance sampling weight to the corresponding log joint posterior density value based on Proposition 1, before computing the estimator and credible set under Dirac delta loss.

Third, Proposition 1 generalizes the results in Inoue and Kilian (2013, 2019) by allowing for the imposition of additional narrative inequality restrictions in sign-identified models (with or

without additional zero restrictions), as discussed in Antolin-Diaz and Rubio-Ramirez (2018). Since narrative restrictions restrict the likelihood, they require the reweighting of the posterior draws based on the importance sampler described in Antolin-Diaz and Rubio-Ramirez (2018). Proposition 1 may be adjusted for the reweighting of the joint posterior density values for each posterior draw as in the case of exclusion restrictions on elements of A_0^{-1}.[16]

3.4 Extensions to partially identified models

The derivations in section 3 are for fully identified structural VAR models. It is conceptually straightforward to extend the analysis under Dirac delta loss to partially identified structural VAR models. This requires the user to integrate out by numerical methods the elements of θ in the joint posterior distribution that are not identified (see Inoue and Kilian 2013). For example, in the model of Uhlig (2005) one would integrate out all responses to shocks other than the monetary policy shock. It should be noted, however, that this Monte Carlo integration approach is much more computationally costly than evaluating the relevant subset of θ under quadratic or absolute loss.

4 Further Extensions

4.1 An Alternative Loss Function

One drawback of the absolute and quadratic loss function is that the Bayes estimator may change, depending on the scaling of the data. This is not of particular concern when all variables are expressed in the same units (say, percentage growth rates), but becomes relevant when some variables

[16] It should be noted that one does not necessarily have to estimate the structural VAR model in the notation used by Arias et al. (2018) and Antolin-Diaz and Rubio-Ramirez (2018) to make use of our results. When using existing code, VAR model estimates can typically be expressed in terms of A_0 and A_+ matrices for the purpose of evaluating the posterior density value of a given posterior draw without having to re-simulate the posterior draws using their notation.

are measured on a different scale than others. Intuitively, under absolute loss or quadratic loss, variables measured on smaller scales will receive less weight in computing the Bayes estimator. This problem does not arise under Dirac delta loss.

A second drawback that affects all conventional loss functions is that the optimal estimator may be sensitive to the scaling of the structural shocks. For example, the scale of the responses and hence the Bayes estimator is affected by whether we compute the response to a one standard deviation shock, as is standard in the structural VAR literature, or we normalize the structural shock to represent a unit shock for one of the model variables. Scaling shocks to represent unit shocks is rare in structural VAR analysis, given that structural shocks affect more than one variable and a shock can be normalized to have a unit impact in one dimension only, but it is desirable for the optimal estimator not to be sensitive to such normalizations.

In this subsection, we consider an alternative loss function that is invariant to both the scale of the data and the scale of the responses. Let A denote a diagonal matrix of the scales of the variables, B the matrix of structural impulse responses at horizon h, and C a diagonal matrix of the magnitudes of the structural shocks underlying the response functions. Then ABC is the matrix of the structural impulse responses at horizon h given these normalizations. Because $vec(ABC) = (C' \otimes A)vec(B)$, the vector of structural impulse responses can be written as $\tilde{\theta} = D\theta$ where θ is the original impulse response vector and D is a diagonal matrix whose diagonal elements are functions of those in A and C.

The objective is for the loss function to be independent of D (and thus of A and B). Let θ_{ij} and $\bar{\theta}_{ij}$ be vectors of the structural responses of the ith variable to the jth structural shock, respectively. Then the cosine of the radius between θ_{ij} and $\bar{\theta}_{ij}$ is given by

$$\sum_{j=1}^{n} \frac{\theta_{ij}\bar{\theta}_{ij}}{\|\theta_{ij}\|\|\bar{\theta}_{ij}\|},$$

and is called the cosine similarity. The cosine similarity does not depend on A or C. Because the cosine similarity is not a distance, we define the *angular loss* by

$$L(\theta, \bar{\theta}) \;=\; \frac{1}{\pi n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \cos^{-1}\left(\frac{\theta_{ij} \bar{\theta}_{ij}}{\|\theta_{ij}\| \|\bar{\theta}_{ij}\|} \right)$$

which is the average of the angular distances between θ_{ij} and $\bar{\theta}_{ij}$ for $i, j = 1, 2..., n$. Minimizing this loss function in expectation with respect to the joint posterior distribution of θ yields the optimal estimator under angular loss. The corresponding $(1-\alpha)100\%$ joint credible set is constructed by retaining the $(1-\alpha)100\%$ draws for θ with the smallest angular loss among the M draws that satisfy the identifying restrictions. The computational cost of this estimator is higher than that of the estimator under absolute loss or quadratic loss, especially when M is very large. As shown in section 5, the use of this loss function tends to generate estimates quite similar to those obtained under conventional loss functions.

4.2 Dealing with Non-Unique Solutions

Whereas the Bayes estimator is unique under absolute and quadratic loss when $n_{irf} \leq n_p$, for $n_{irf} > n_p$ there may be situations in which more than one posterior draw minimizes the loss function in expectation. More generally, under Dirac delta loss and angular loss, the Bayes estimator need not be unique, even when $n_{irf} \leq n_p$. This does not affect the construction of the joint credible set, but calls for a change in the construction of the Bayes estimator. Recall that the infeasible version of our Bayes estimator is

$$\Theta_0 \;=\; \{\bar{\theta} \in \Theta :\; E_\theta[L(\theta, \bar{\theta})] \;\leq\; E_\theta[L(\theta, \tilde{\theta})] \text{ for } \forall\, \tilde{\theta} \in \Theta\} \tag{36}$$

where the expectations are with respect to the joint posterior distribution of θ. When $n_{irf} > n_p$, Θ_0 is not necessarily a singleton. Since it is difficult to determine a priori whether the solution is unique or not, it is useful to have a diagnostic tool for assessing this question. A practical solution to this problem inspired by Chernozhukov, Hong and Tamer's (2007), is to first estimate

$$\widehat{\bar{\theta}}_M \;=\; \text{argmin}_{\bar{\theta}\in\widehat{\Theta}_M} \frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)},\bar{\theta}), \tag{37}$$

where $\widehat{\Theta}_M$ consists of M posterior draws and $\theta^{(i)}$ denotes the ith posterior draw. Then the estimate of the set Θ_0, for a given loss function, is given by

$$\widehat{\Theta}_{0,M} \;=\; \{\overline{\theta}\in\widehat{\Theta}_M : 100\times\left(\frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)},\bar{\theta}) - \frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)},\widehat{\bar{\theta}}_M)\right) \Big/ \frac{1}{M}\sum_{i=1}^{M} L(\theta^{(i)},\widehat{\bar{\theta}}_M) \leq c_M\} \tag{38}$$

where c_M is a sequence of positive numbers that satisfies $c_M \;\to\; 0$ and $1/(c_M\sqrt{M}) = o(1)$ as $M \;\to\infty$. In practice, we set $c_M = \log(M)/\sqrt{M}$, which satisfies these conditions. In short, the approach is to include all posterior draws in the c_M%-neighborhood of the minimum expected loss. As $M\to\infty$, this neighborhood is shrunk to zero, so in the limit all remaining solutions have exactly the same expected loss. As the empirical examples in the next section shows, in many cases $\widehat{\Theta}_0$ consists of a singleton, and, when not, there typically are only two solutions in this set that are quite similar.

5 Empirical illustrations

In this section, we present two empirical VAR examples based on diffuse Gaussian-inverse Wishart priors for the reduced-form parameters. All models include an intercept.

5.1 Example 1

Our first empirical example is a stylized model of the U.S. macroeconomy based on short-run and long-run exclusion restrictions, as discussed in Rubio-Ramirez, Waggoner and Zha (2010). This specific example was chosen because models combining short-run and long-run restrictions are not covered by the theoretical results in Inoue and Kilian (2013, 2019). We specify a quarterly VAR(4) model for real GNP growth ($\Delta \log Y$), the federal funds rate (R) and GNP deflator inflation ($\Delta \log P$). Variation in these data is explained in terms of an aggregate demand shock, an aggregate supply shock, and a monetary policy shock. The identifying restrictions are that aggregate demand shocks have no long-run effect on real output and that monetary policy shocks have neither an impact effect on real output nor a long-run effect on real output. The model is exactly identified. The estimation period is 1954.IV-2007.IV. We solve for the structural impact multiplier matrix, as discussed in Rubio-Ramirez, Waggoner and Zha (2010). The sign of the monetary policy shock is normalized to imply an increase in the federal funds rate on impact. The sign of the aggregate supply shock is normalized to ensure that the aggregate supply shock raises real output on impact. Given that the structural model is only loosely restricted, we follow the convention of reporting 68% credible sets. The maximum impulse response horizon is 12 quarters. The number of estimated structural impulse responses, n_{irf}, exceeds the number of estimated structural parameters, n_p.

Although we are considering all 9 structural impulse functions when minimizing the expected loss, in the interest of space, only a subset of the results are shown. Figure 2 focuses on the response of the model variables to an unanticipated monetary tightening. Such a shock would be expected to raise the interest rate, lower real GNP and lower GNP deflator inflation. We report estimators and 68% joint credible regions under each of the four loss functions considered in this paper. For comparison we also report the conventional posterior median response function and posterior mean response function in the first two columns together with the conventional pointwise quantile error

bands. All results shown are based on $M = 5,000$ posterior draws.

Our optimal estimates of θ allow for multiple solutions. As Figure 2 shows, only under quadratic loss, the optimal estimator is not unique, but the two optimal solutions found by the algorithm are very similar. The differences in the impulse response estimates across loss functions are comparatively minor. All estimates indicate a transitory increase in the interest rate and a persistent decline in real GNP. Inflation is not very responsive. There is no indication that using the angular loss function rather than conventional loss functions makes much of a difference for the estimates. Angular and absolute loss imply somewhat tighter joint credible regions than quadratic or Dirac delta loss, as measured by the range of responses. Although in this example the difference between the Bayes estimator under absolute and under quadratic loss and the posterior median and mean response functions, respectively, is small, Figure 2 confirms that the use of pointwise error bands causes the user to understate the estimation uncertainty at short as well as long horizons. In some cases, the range spanned by the responses in the joint credible set is three times as wide.

One of the advantages of lowest posterior risk joint credible sets is that we retain information about the shape of the response functions contained in the joint credible set. One way of highlighting this information is to represent the elements of the joint credible region in different colors differentiated by probability content. Figure 3 demonstrates this approach based on the estimates obtained under absolute loss.

Even more useful in practice is the ability to highlight in different colors the relationship across the responses of different model variables. Figure 4, for example, shows in black the subset of posterior draws that is consistent with a persistent decline in deflator inflation in response to an unexpected monetary tightening. Figure 4 shows not only that the credible set for the output responses becomes tighter under this restriction, but posterior draws implying very large interest rate responses are systematically ruled out. Such information would have been irretrievably lost, if

we had simply constructed a joint error band by constructing an envelope around the joint credible set. Other examples have been discussed in Inoue and Kilian (2016). For example, the ability to plot individual posterior draws for the responses of inflation and output is extremely valuable in assessing whether oil price shocks are stagflationary for the U.S. economy. It may also be used to assess whether response functions are hump-shaped or whether there is delayed overshooting, for example.

5.2 Example 2

The second empirical example is a monthly VAR(24) model of the global market for crude oil based on Kilian and Murphy (2014). The model variables include the percent change in global crude oil production, an appropriate measure of the global business cycle, the log real price of oil, and the change in global crude oil inventories. The model differentiates between various shocks to the demand for oil including flow demand and storage demand and shocks to the flow supply of oil. The structural shocks in the oil market block are identified based on static and dynamic sign restrictions, bounds on the impact price elasticities of demand and supply, and narrative restrictions. Unlike in the earlier example, the structural impulse responses are only set-identified. The estimation period is 1973.2-2018.6. Details of the construction of the data, of the elasticity bounds, and of the specification of the narrative restrictions can be found in Zhou (2020).

The model is estimated using a conventional uniform-Gaussian inverse Wishart prior, as described in section 3. Because the model is estimated subject to narrative restrictions on the historical decomposition of the real price of oil, the posterior draws have to be reweighted based on an importance sampler, as described in Antolin-Diaz and Rubio-Ramirez (2018). The maximum impulse response horizon, H, is 17 months. Thus, the number of estimated structural parameters, $n_p = n^2(p+1)$, exceeds the number of estimated structural impulse responses, $n_{irf} = n^2(H+1)$.

As in the previous example, we evaluate the model based on all 16 impulse response functions jointly, but, for illustrative purposes, we report only a subset of the results. Figure 5 reports the Bayes estimates under each loss function as well the corresponding 68% joint credible sets based on $M = 10,000$ posterior draws. It also reports the conventional pointwise estimates and 68% pointwise quantile error bands for comparison.

Figure 5 focuses on the response of the real price of oil to selected oil demand and oil supply shocks. All shocks have been normalized to imply an increase in the real price of oil on impact. All Bayes estimates are unique and, notwithstanding some minor differences, the estimates are broadly similar across loss functions. A disruption to the flow of oil supplies causes a modest, but persistent increase in the real price of oil that declines over time. Positive flow demand shocks are associated with a larger and more persistent increase in the real price of oil. Except under quadratic loss, this response peaks after three or four months. Storage demand shocks have an even larger effect on the real price of oil, but that response declines more quickly than in the case of the flow demand shock. The responses to the flow demand and storage demand shocks are precisely estimated in that the joint credible region does not include the zero line at all but the largest horizons. There is not much difference in the width of the credible sets across loss functions, as measured by the range of the responses. As in the previous example, the corresponding pointwise 68% error bands are much too narrow. In some cases, they do not even contain the Bayes estimate.

6 Concluding remarks

Building on insights in Bernardo (2010), we derived the Bayes estimator of vectors of structural impulse responses and the corresponding lowest posterior risk joint credible sets under four alternative loss functions. We also generalized the Bayes estimator to allow for the possibility that the estimator may not be unique.

Our analysis shows that under absolute and quadratic loss, respectively, posterior median and mean response functions are not in general the Bayes estimator of the vector θ of structural impulse responses obtained by stacking the impulse responses of interest. Specifically, in the empirically more relevant case of $n_{irf} > n_p$, these pointwise estimators will not coincide with the Bayes estimator of θ, even as the number of posterior draws approaches infinity. In contrast, in the empirically less relevant case of $n_{irf} \leq n_p$. they will recover the Bayes estimator in the limit, although these estimators are likely to differ in practice even when the number of posterior draws is large.

Whether $n_{irf} > n_p$ or $n_{irf} \leq n_p$, conventional pointwise impulse response error bands for θ are invalid. We showed that pointwise error namds tend to be much too narrow. One advantage of our approach of computing valid joint credible sets for θ under absolute and quadratic loss, respectively, is that it generates as a by-product the Bayes estimator. Another advantage is that it provides more information about likely departures from the Bayes estimator than joint error bands. We showed how visual aids may be used to examine the shape of the response functions as well as the comovement across structural response functions.

We also derived the Bayes estimator and the lowest posterior risk joint credible set under Dirac delta, greatly extending the range of applications that can be handled based on this loss function, which has received increasing attention in the recent literature. Finally, we considered a new loss function, referred to as angular loss, that renders the Bayes estimator of θ robust to the scaling of the data and of the structural shocks.

Our empirical evidence suggests that there is little to choose between the absolute, quadratic and Dirac delta loss based on computational cost in typical applications (except that, in partially identified models, Dirac delta loss requires the numerical derivation of the marginalized distribution). The Bayes estimator based on angular loss is more computationally costly when the number of posterior draws becomes very large. Our two empirical illustrations suggest that the four Bayes

estimators proposed in this paper tend to generate broadly similar results.

Our theoretical analysis relied on the standard Bayesian approach to estimating structural VAR models as discussed in Rubio-Ramirez, Waggoner and Zha (2010), Arias et al. (2018) and Antolin-Diaz and Rubio-Ramirez (2018). The Bayes estimators and joint credible sets we derived under additively separable quadratic and absolute loss may also be directly applied to the alternative Bayesian approaches discussed in Baumeister and Hamilton (2018) and Plagborg-Møller (2019). We leave for future research the question of how to derive analogous results for the joint posterior density of the structural impulse responses under their assumptions, as required for the implementation of the Bayes estimator under Dirac delta loss.

Although we focus on VAR impulse response analysis, it should be noted that the same techniques for estimation and joint inference may also be adapted to other vector valued VAR statistics such as forecast error variance decompositions and historical decompositions or the estimation of sequences of structural shocks.

Beyond structural VAR modeling, our approach could be easily adapted to estimate the path of a forecast and to characterize the joint uncertainty about this path. For example, during economic crises policymakers are routinely concerned with forecasting the path of the recovery over the next year or two. They are not interested in forecasting economic outcomes in a specific quarter in isolation. Under absolute or quadratic loss, the methods we developed for impulse response analysis may be directly applied to a panel of M posterior draws of economic forecasts across multiple horizons to recover a Bayes estimate of the forecast path and of the associated joint credible region.

Our approach may also be used in the evaluation of impulse responses based on Bayesian estimates of DSGE models. For example, Herbst and Schorfheide (2015) report posterior mean response functions and pointwise quantile error bands computed based on draws from the posterior

distribution of the structural DSGE model parameters. Such estimates are subject to the same concerns as posterior mean response functions in structural VAR models. The posterior mean response function may generate estimates inconsistent with the DSGE model structure, if the number of responses exceeds the number of structural DSGE model parameters, and the pointwise error bands by construction misrepresent the uncertainty about the response functions. It would be straightforward to employ the tools we developed under quadratic loss or absolute loss to address this concern. Extensions to Dirac delta loss would be more involved and would likely involve numerical differentiation.

Finally, our tools can be adapted to formally characterize the joint prior of the impulse responses in sign-identified VAR models and to examine how sensitive the posterior is to this prior. This question has recently been at the forefront of the debate about the use of conventional priors in estimating sign-identified VAR models and the reliability of the resulting posterior estimates (see Inoue and Kilian 2020).

Acknowledgments

The views expressed in the paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. We thank Juan Antolin-Diaz, Ed Herbst, Mikkel Plagborg-Møller, Helmut Lütkepohl and Benjamin Wong, the associate editor, two anonymous referees, and the editor for helpful comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Aastveit, K.A, Bjørnland, H.C., Thorsrud, L.A., 2015. What drives oil prices? Emerging vs. developed economies. Journal of Applied Econometrics 30, 1013-1028.

2. Antolin-Diaz, J., Rubio-Ramirez, J.F., 2018. Narrative sign restrictions for SVARs. American Economic Review 108, 2802-2839.

3. Arias, J., Rubio-Ramirez, J.F., Waggoner, D.F., 2018. Inference based on SVARs identified with sign and zero restrictions: Theory and applications. Econometrica 86, 685-720.

4. Baumeister, C., Hamilton, J.D., 2018. Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations. Journal of Monetary Economics 100, 48-65.

5. Berger, J.O., 1985. Statistical decision theory and Bayesian analysis. 2nd ed. Springer, New York.

6. Bernardo, J.M., 2010. Integrated objective Bayesian estimation and hypothesis testing. Bayesian Statistics 9, 1-25.

7. Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. John Wiley and Sons, Chichester, UK.

8. Blanchard, O.J., 1989. A traditional interpretation of macroeconomic fluctuations. American Economic Review 79, 1146–1164.

9. Bruder, S., Wolf, M., 2018. Balanced bootstrap joint confidence bands for structural impulse response functions. Journal of Time Series Analysis 39, 641-664.

10. Chen, X., Christensen, T.M., Tamer, E., 2018. Monte Carlo confidence sets for identified sets. Econometrica 86, 1965-2018.

11. Cheng, X., Han, X., Inoue, A., 2020. Instrumental variable estimation of structural VAR models robust to possible non-stationarity. Manuscript, Vanderbilt University.

12. Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243-1284.

13. Cochrane, J.H., 1994. Shocks. Carnegie-Rochester Conference Series on Public Policy, 41, 295-364.

14. Cross, J.L., Nguyen, B.H., Tran, T.D, 2019. The role of precautionary and speculative demand in the global market for crude oil. Manuscript, BI Norwegian Business School.

15. Cushman, D.O., Zha, T., 1997. Identifying monetary policy in a small open economy under flexible exchange rates. Journal of Monetary Economics 39, 433-448.

16. Eichenbaum, M., Evans, C.L. 1995. Some empirical evidence on the effects of monetary policy shocks on exchange rates. Quarterly Journal of Economics 110, 975-1009.

17. Fry, R., Pagan, A.R., 2011. Sign restrictions in structural vector autoregressions: A critical review. Journal of Economic Literature 49, 938-960.

18. Giacomini, R., Kitagawa, T., (2020). Robust Bayesian inference for set-identified models. Econometrica, forthcoming.

19. Herbst, E., Schorfheide, F., 2015. Bayesian estimation of DSGE models. Princeton University Press, Princeton.

20. Herrera, A.M., Rangaraju, S.K., 2020. The effects of oil supply shocks on U.S. economic activity: What have we learned? Journal of Applied Econometrics, 35, 141-159.

21. Herwartz, H., Plödt, M., 2016. The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. Journal of International Money and Finance 61, 30-44.

22. Inoue, A., Kilian, L., 2013. Inference on impulse response functions in structural VAR models. Journal of Econometrics 177, 1-13.

23. Inoue, A., Kilian, L., 2016. Joint confidence sets for structural impulse responses. Journal of Econometrics 192, 421-432.

24. Inoue, A., Kilian, L. 2019. Corrigendum to 'Inference on impulse response functions in structural VAR models' [J. Econometrics 177 (2013), 1-13]. Journal of Econometrics 209, 139-143.

25. Inoue, A., Kilian, L. 2020. The Role of the Prior in Estimating VAR Models with Sign Restrictions. Manuscript, Federal Reserve Bank of Dallas.

26. Kilian, L., Lütkepohl, H., 2017. Structural vector autoregressive analysis. Cambridge University Press, New York.

27. Kilian, L., Murphy, D.P., 2012. Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models. Journal of the European Economic Association 10, 1166-1188.

28. Kilian, L., Murphy, D.P., 2014. The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Econometrics 29, 454-478.

29. Kilian, L., Zhou, X., 2020a. Oil prices, exchange rates and interest rates. Manuscript, Federal Reserve Bank of Dallas.

30. Kilian, L., Zhou, X., 2020b. Does drawing down the U.S. strategic petroleum reserve help stabilize oil prices? Journal of Applied Econometrics, forthcoming.

31. Koop, G., 1996. Parameter uncertainty and impulse response analysis. Journal of Econometrics 72, 135-149.

32. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2015a. Confidence bands for impulse responses: Bonferroni vs. Wald. Oxford Bulletin of Economics and Statistics 77, 800-821.

33. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2015b. Comparison of methods for constructing joint confidence bands for impulse response functions. International Journal of Forecasting 31, 782-798.

34. Lütkepohl, H., Staszewska-Bystrova, A., Winker, P., 2018. Calculating joint confidence bands for impulse response functions using highest density regions. Empirical Economics 55, 1389-1411.

35. Montiel Olea, J.L., Plagborg-Møller, M., 2019. Simultaneous confidence bands: Theory, implementation, and an application to SVARs. Journal of Applied Econometrics 34, 1-17.

36. Mumtaz, H., Surico, P., 2009. The transmission of international shocks: A factor-augmented VAR approach. Journal of Money, Credit and Banking 41, 71-100.

37. Nam, D.,Wang, J., 2018. Mood swings and business cycles: Evidence from sign restrictions. Journal of Money, Credit, and Banking 51, 1623-1649.

38. Plagborg-Møller, M., 2019. Bayesian inference on structural impulse response functions. Quantitative Economics 10, 145-184.

39. Rausser, G., and Stürmer, M., 2020. A dynamic analysis of collusive action: The case of the world copper market, 1882-2016. Manuscript, Federal Reserve Bank of Dallas.

40. Rubio-Ramirez, J.F., Waggoner, D., Zha, Z., 2010. Structural vector autoregressions: Theory of identification and algorithms for inference. Review of Economic Studies 77, 665-696.

41. Sims, C.A., Zha, T., 1999. Error bands for impulse responses. Econometrica 67, 1113-1156.

42. Uhlig, H. 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics 52, 381-419.

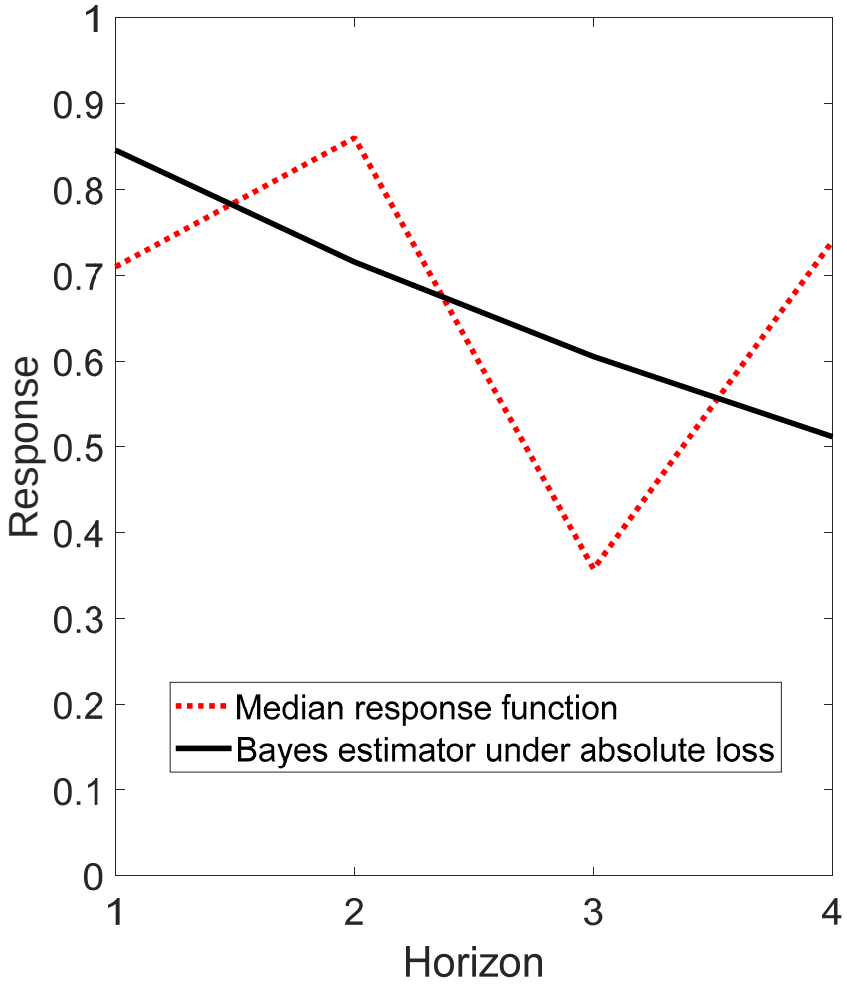
43. Waggoner, D., Zha, T., 2012. Confronting model misspecification in macroeconomics. Journal of Econometrics 171, 167-184.

44. Woodford, M., 2003. Interest and prices: Foundations of a theory of monetary policy. Princeton University Press, Princeton.

45. Zha, T., 1999. Block recursion and structural vector autoregressions. Journal of Econometrics 90, 291-316.

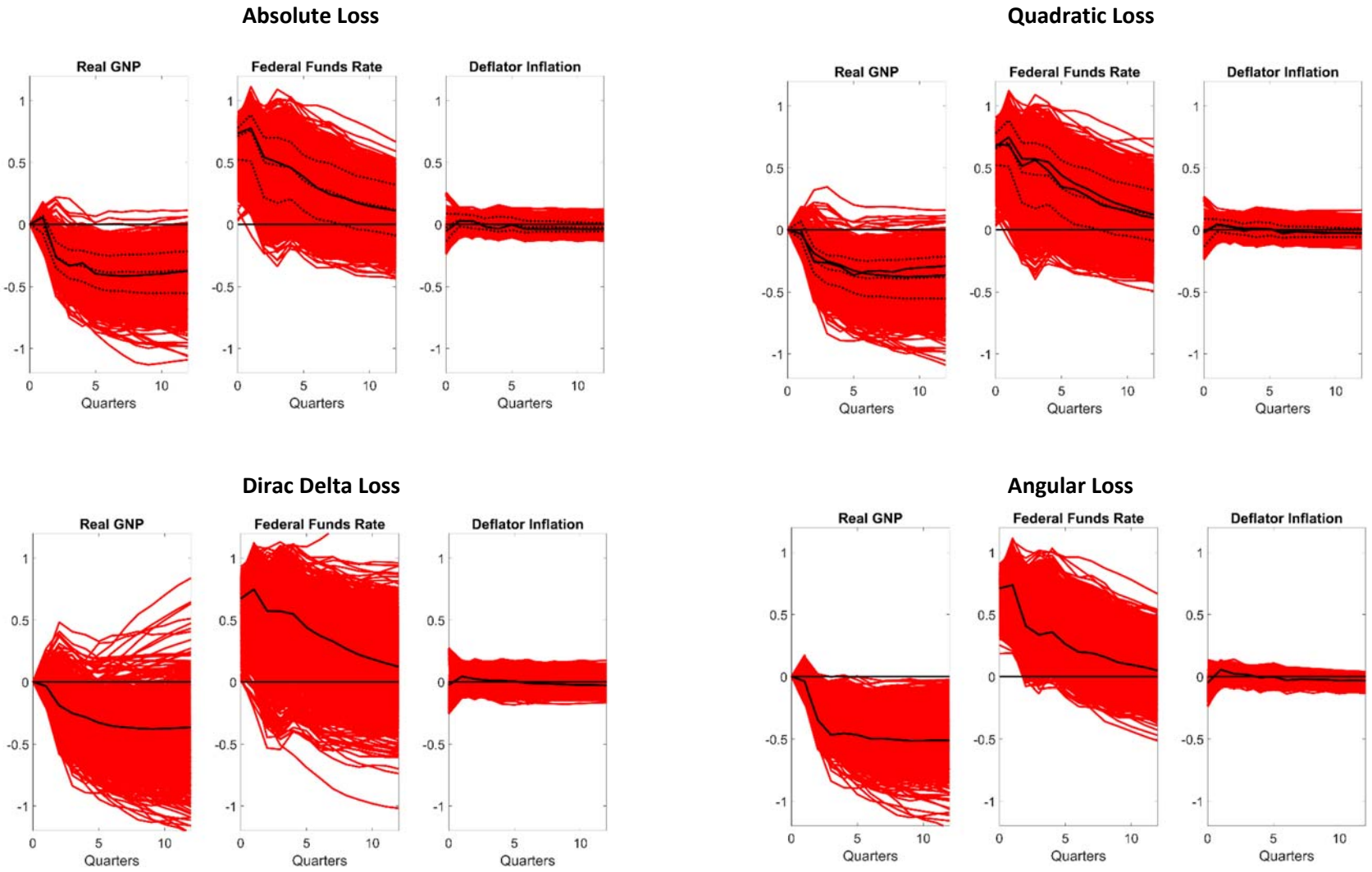
46. Zhou, X., 2020. Refining the workhorse oil market model," Journal of Applied Econometrics, 35, 130-140.

Figure 1: AR(1) example of how median response functions may distort the shape of the impulse response function



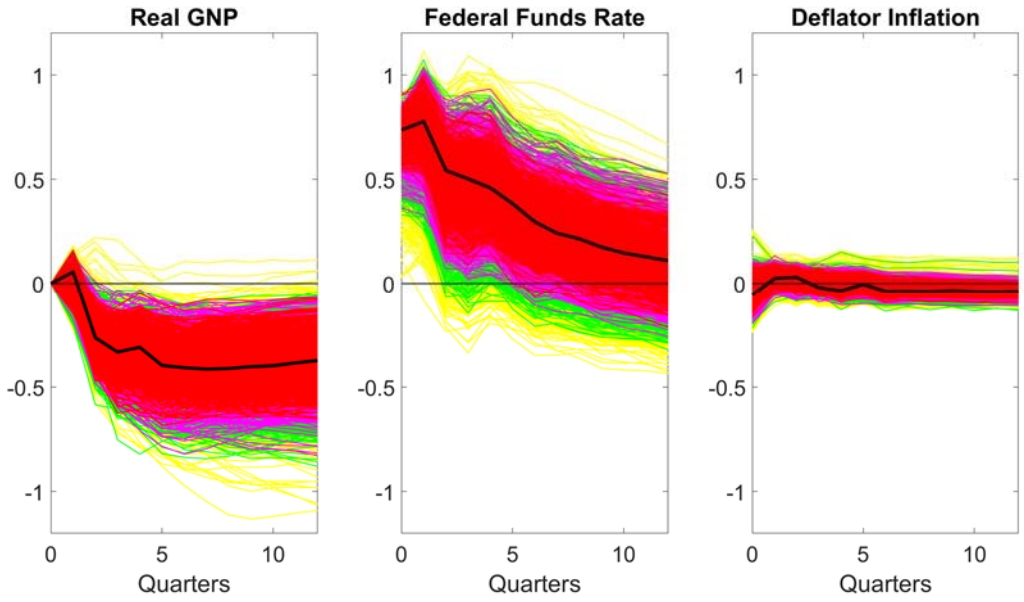
NOTES: Responses generated from slope parameter in AR(1) model with posterior $\rho \sim (\rho_{mean}, 1/n_T)$ with $\rho_{mean} = 0.7$ and $n_T = 0.769$. All estimates are based on 10,000 random draws from the posterior.

Figure 2: Responses to exogenous U.S. real interest rate shock: 68% joint credible set under alternative loss functions with Bayes estimates highlighted in black



NOTES: Pointwise estimates, when available, are shown as dotted lines for comparison.

**Figure 3: Responses to exogenous U.S. real interest rate shock:
Bayes estimate under absolute loss and
joint credible regions with various probability contents**



NOTES: Red: 17%. Magenta: 34% Green: 51% Yellow: 68% probability content.

**Figure 4: Responses to exogenous U.S. real interest rate shock:
68% joint credible set under absolute loss with posterior
draws that generate a persistent decline in inflation highlighted in black**

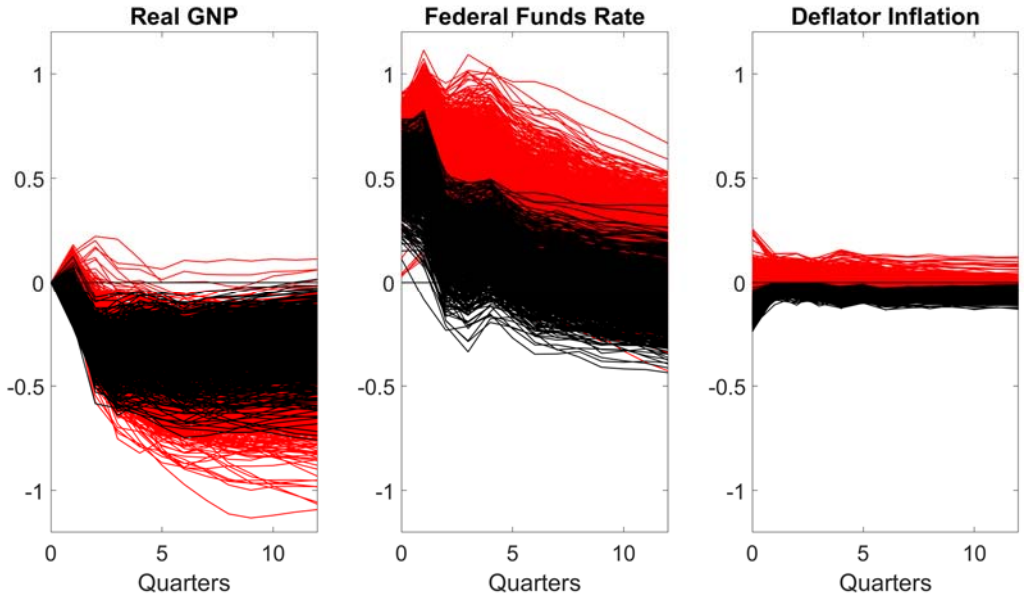
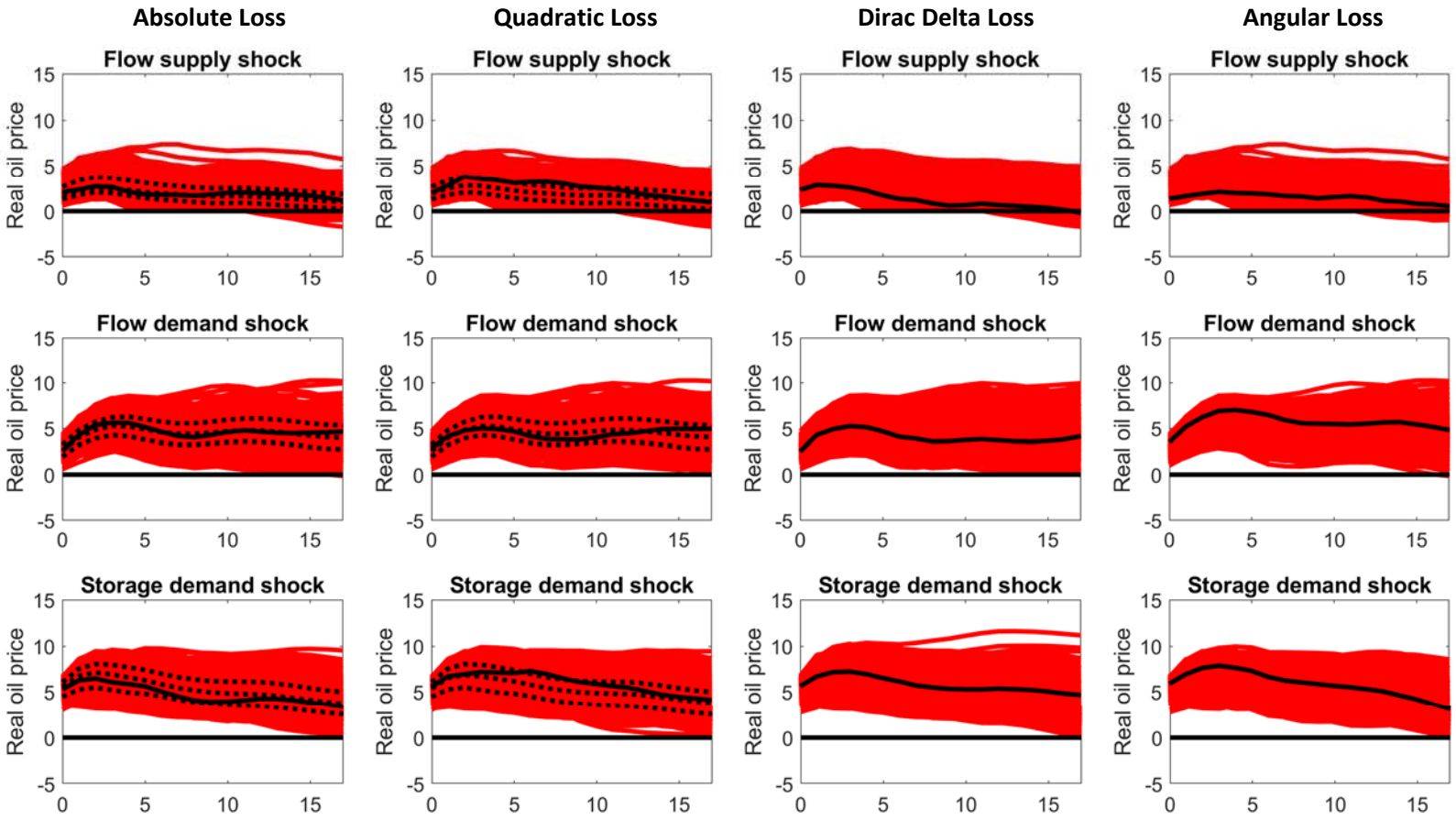


Figure 5: Responses of the Real Price of Oil:
68% joint credible set under alternative loss functions with Bayes estimates highlighted in black



NOTES: Pointwise estimates, when available, are shown as dotted lines for comparison.