

Estimating Posterior Sensitivities with Application to Structural Analysis of Bayesian Vector Autoregressions

Liana Jacobi, Dan Zhu & Mark Joshi

To cite this article: Liana Jacobi, Dan Zhu & Mark Joshi (2025) Estimating Posterior Sensitivities with Application to Structural Analysis of Bayesian Vector Autoregressions, Journal of Business & Economic Statistics, 43:1, 134-149, DOI: [10.1080/07350015.2024.2329639](https://doi.org/10.1080/07350015.2024.2329639)

To link to this article: <https://doi.org/10.1080/07350015.2024.2329639>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 22 Apr 2024.



[Submit your article to this journal](#)



Article views: 977



[View related articles](#)



[View Crossmark data](#)

Estimating Posterior Sensitivities with Application to Structural Analysis of Bayesian Vector Autoregressions

Liana Jacobi^a, Dan Zhu^b, and Mark Joshi^{*a} 

^aDepartment of Economics, University of Melbourne, Parkville, VIC, Australia; ^bDepartment of Business Statistics and Econometrics, Monash University, Clayton, VIC, Australia

ABSTRACT

The inherent feature of Bayesian empirical analysis is the dependence of posterior inference on prior parameters, which researchers typically specify. However, quantifying the magnitude of this dependence remains difficult. This article extends Infinitesimal Perturbation Analysis, widely used in classical simulation, to compute asymptotically unbiased and consistent sensitivities of posterior statistics with respect to prior parameters from Markov chain Monte Carlo inference via Gibbs sampling. The method demonstrates the possibility of efficiently computing the complete set of prior sensitivities for a wide range of posterior statistics, alongside the estimation algorithm using Automatic Differentiation. The method's application is exemplified in Bayesian Vector Autoregression analysis of fiscal policy in U.S. macroeconomic time series data. The analysis assesses the sensitivities of posterior estimates, including the Impulse response functions and Forecast error variance decompositions, to prior parameters under common Minnesota shrinkage priors. The findings illuminate the significant and intricate influence of prior specification on the posterior distribution. This effect is particularly notable in crucial posterior statistics, such as the substantial absolute eigenvalue of the companion matrix, ultimately shaping the structural analysis.

ARTICLE HISTORY

Received July 2022
Accepted March 2024

KEYWORDS

Bayesian vector autoregression; Sensitivity analysis; Structural analysis; Shrinkage priors; Time series analysis

1. Introduction

Empirical analysis through Bayesian inference relies on posterior distributions that combine information from the data using the model likelihood and prior information on model parameters. The latter is characterized by prior distributions that depend on a set of hyperparameters (prior parameters). As prior parametric families are usually determined based on computational considerations, researchers are mainly required to specify the prior parameters.

In practice, researchers adopt various strategies to set these prior parameters, depending on available data, modeling context, identification issues, or existing prior information. These strategies encompass the use of perceived uninformative default priors and more informative priors established through different methods of prior parameter elicitation (Giannone, Lenza, and Primiceri 2015; Baumeister and Hamilton 2019; Jarociński and Marcet 2019; Jacobi et al. 2021). One domain that experiences an active debate regarding prior parameter specification is the Bayesian vector autoregressive (BVAR) analysis with shrinkage priors for macroeconomic forecasting and structural analysis (Amir-Ahmadi, Matthes, and Wang 2018; Huber and Feldkircher 2019; Baumeister, Korobilis, and Lee 2020).



In formal terms, the dependence of prior parameters locally can be captured by observing changes in posterior inference resulting from modifications in any of the prior parameters, as formalized by the sensitivity of a statistic (Berger, Insua, and Ruggeri 2000; Gustafson 2000). However, calculating the exact

derivative of a posterior statistic with respect to even a single hyperparameter is typically analytically intractable, given that posterior distributions manifest as high-dimensional integrals requiring numerical solutions. Efficient Markov chain Monte Carlo (MCMC) simulation methods, especially Gibbs sampling, offer popular tools for constructing transition kernels to simulate posterior distributions and estimate posterior statistics across various modeling contexts.¹ Constrained by the computational burden, empirical studies often resort to a less formal sensitivity analysis by re-executing the Gibbs algorithm under alternative prior assumptions (Del Negro and Schorfheide 2008; Jarociński and Marcet 2019; Jacobi et al. 2021).


In this article, we present an approach for sensitivity analysis of MCMC inference. This approach is based on extending the Infinitesimal Perturbation Analysis of the simulation path, traditionally employed in independent samples, to dependent MCMC samples. In particular, we introduce IPA analysis for MCMC inference in Gibbs sampling and leverage its dependence on smooth transition kernels. Notably, Gibbs sampling is widely favored in empirical research due to its advantageous computational properties (Greenberg 2012), and its broad applicability.² When combined with suitable conjugate priors, data

¹Throughout this article, the term “posterior statistic” (or “posterior output”) pertains to the posterior mean associated with a certain function of parameters.

²This includes numerous implementations in various software packages, such as R-packages like Bayesm by Peter Rossi for Marketing and Microeconometrics, MCMCpack, and the BUGS packages.

CONTACT Dan Zhu  dan.zhu@monash.edu  Department of Business Statistics and Econometrics, Monash University, Clayton, VIC, 3168, Australia.

*Mark Joshi passed away on the 8th of Oct, 2017. He was Professor at the Department of Economics at the University of Melbourne

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

© 2024 The Authors. Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

augmentation techniques, and smoothing methods (Albert and Chib 1993; Frühwirth-Schnatter 1994), Gibbs-based inference can be extended to complex model settings, encompassing state-space models (Durbin and Koopman 2002), treatment effects models (Jacobi, Wagner, and Frühwirth-Schnatter 2016), multivariate demand analysis (Briggs et al. 2017), and time series models like Bayesian Vector Autoregressions (BVARs) (Koop and Korobilis 2010; Camacho, Gadea, and Gómez-Loscos 2020).

To establish the theoretical properties of our proposed IPA derivative estimator, we break down the Gibbs algorithm into three distinct categories of mappings: (a) the computation of hyperparameters for the transition kernel, (b) the generation of random variables using the transition kernel, and (c) the conversion of parameter samples into posterior statistics, encompassing measures like the mean, covariance, and more intricate outcomes like forecasts and impulse response functions. We establish the IPA derivative by employing a chain-rule argument and expand the theoretical IPA results from the classical simulation context to MCMC under a set of regularity conditions. In particular, we prove that the resulting estimator exhibits desirable asymptotic properties of unbiasedness and consistency. For precise implementation, we employ tools from Automatic Differentiation (AD) (as discussed in (Griewank and Walther 2008)) to facilitate derivative computation in parallel with the original MCMC estimation algorithm.

The Likelihood Ratio (LR) method is frequently employed as an unbiased alternative for IPA derivative estimation in classical simulation. It involves shifting the parameters of interest into the likelihood function to compute derivative outcomes. Within the MCMC framework, Pérez, Martin, and Rufo (2006) introduced an LR approach for inference within the exponential parametric family that relies on the analytical partial derivative of the score function. Thus, the application of the LR approach to assess sensitivities from posterior to prior has far been confined to computing posterior mean sensitivities concerning prior coefficient means (Müller 2012). Another alternative to the infinitesimal perturbation in the IPA approach is finite perturbation via the finite-difference (FD) method. While widely applicable, the FD method lacks the unbiasedness property of IPA and LR, and its effectiveness depends on the chosen bump size. In principle, a smaller bump size should lead to a reduced bias under a set of regularity conditions, but this is often unattainable in practice due to numerical instability when dealing with extremely small numbers in computer operations (Glasserman 2013).

We demonstrate the unbiasedness and precision inherent in the IPA-based derivative computation through simulation experiments conducted in a standard small Bayesian Vector Autoregression (BVAR) setup. Our benchmarking study concentrates on assessing the sensitivities of the posterior means and variances of the regression coefficients. To maintain consistency with the simulation literature, we incorporate the LR method whenever feasible as an alternative for unbiased derivative estimation. For the remaining comparisons, we rely on FD estimates. Our simulation results align with our theoretical conclusions and substantiate classical simulation literature findings (Glasserman 2013). They reveal that IPA estimates are not only unbiased when compared to LR estimates but also boast higher precision. Furthermore, in contrast to the biased yet general FD approach, IPA sensitivity estimates exhibit computational

efficiency and stability. In contrast, the FD estimates show considerable dependence on the chosen bump size. Within the same BVAR framework, we also expand the application of IPA sensitivities to encompass the Impulse response function (IRF), Forecast error variance decomposition (FEVD), and the eigenvalues of the companion matrix for model stability assessment. Through this extension, we showcase the notable influence that shrinkage priors exert on these posterior statistics when juxtaposed with weakly informative priors.

In our application, we consider a standard fiscal policy Bayesian Vector Autoregression (BVAR) model encompassing government spending, GDP, and tax revenue. This framework follows the spirit of Ramey (2019) and employs a widely adopted Minnesota-style shrinkage prior, rooted in the popular Litterman prior (Doan, Litterman, and Sims 1984; Litterman 1986). Our findings suggest that the sensitivities of posterior inference primarily revolve around the prior means and variances pertaining to the GDP equation. This is evident from the sensitivities of prior parameters observed in posterior parameter estimates, and these sensitivities subsequently carry through to influence the IRFs and FEVDs. Focusing on the government spending shock of the log GDP, we find that the impact of prior mean is less than that of the prior variance specifications on the IRF across the complete forecast horizon. Results obtained by rerunning the MCMC (at a higher computational cost) confirm the results of the first-order approximation of the IRFs constructed by our IPA derivatives. Overall, the application showcases that our method provides an additional tool for practitioners using the BVAR approach.

The remainder of the article is organized as follows. [Section 2](#) introduces our proposed framework for IPA derivative analysis of MCMC output and provides its asymptotic convergence properties for Gibbs-based MCMC. In [Section 3](#), we discuss implementation aspects and present a simulation study to compare the proposed IPA-derivatives via AD against two alternatives derivative estimation methods. [Section 4](#) illustrates IPA-derivative analysis in the context of a fiscal policy VAR. [Section 5](#) concludes.

2. IPA based Sensitivity Analysis for Gibbs Inference

Bayesian Inference through Gibbs samplers stands as a fundamental tool in statistical and econometric inference. Over time, Gibbs simulation algorithms have retained their popularity as a preferred method for Bayesian statistical analysis due to their favorable computational attributes and wide-ranging applicability. In this section, we present a novel approach designed to tackle the methodological and computational hurdles inherent in evaluating the robustness of posterior statistics. Our approach focuses on quantifying sensitivities with respect to the prior hyperparameters, as introduced in the realm of Bayesian robustness literature by Berger, Insua, and Ruggeri (2000) and Gustafson (2000). Rather than relying on methods like rerunning the MCMC estimation (numerical finite differencing) or relying on analytical expressions (symbolic differentiation, likelihood ratio method) to evaluate the sensitivities of prior parameters, we leverage the concept of IPA. This approach involves monitoring the undisturbed trajectory of the MCMC simulation

as it evolves, allowing us to predict the effects of small changes in input parameters (as they tend toward zero) on both parameter draws and associated statistics.

2.1. Setting

Consider the setting where some data \mathbf{y} is generated via a data generating process that depends on a vector of parameters $\boldsymbol{\theta} \in \Omega \subset \mathbf{R}^{\dim(\boldsymbol{\theta})}$. Bayesian inference about the model parameters $\boldsymbol{\theta}$ is based on the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\eta}_0, \mathbf{y})$ that combines the model likelihood $f(\mathbf{y}|\boldsymbol{\theta})$ given the model parameters and their prior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\eta}_0)$ where the latter depends on a set of hyperparameters $\boldsymbol{\eta}_0 \in \mathbb{S}^\eta \subset \mathbf{R}^{\dim(\boldsymbol{\eta})}$. In practice, MCMC analysis provides an efficient and popular tool to simulate the posterior distribution, which is typically a high dimensional integral and not analytically tractable.

MCMC-based posterior inference relies on the construction of a MCMC chain of parameter draws, $\{\boldsymbol{\theta}^g\}$, with a Markov transition kernel $\pi(\cdot|\boldsymbol{\theta}^{g-1}, \boldsymbol{\eta}_0, \mathbf{y})$ at each iteration g such that

$$\boldsymbol{\theta}^g|\boldsymbol{\theta}^{g-1} \sim \pi(\cdot|\boldsymbol{\theta}^{g-1}, \boldsymbol{\eta}_0, \mathbf{y}) \text{ for } g = 1, 2, \dots \quad (1)$$

given chain starting values $\boldsymbol{\theta}^0$, $\boldsymbol{\theta}^g|\boldsymbol{\theta}^0 \rightarrow \pi(\boldsymbol{\theta}|\boldsymbol{\eta}_0, \mathbf{y})$ converges to the posterior distribution of interest. Bayesian inference aims to obtain estimates of a set of posterior statistics of interest, $S \in \mathbb{S}^S \subset \mathbf{R}^{\dim(S)}$

$$\alpha(\boldsymbol{\eta}_0, \mathbf{y}) = \mathbb{E}_\pi[S(\boldsymbol{\theta})|\boldsymbol{\eta}_0, \mathbf{y}] \quad (2)$$

that include estimates of first and second order moments, as well as interval estimates which will also depend on the prior parameter specified for the inference. Our approach encompasses more complex posterior statistics, including IRF estimates as illustrated in the real data analysis.

2.2. Prior Parameter Dependence via Local Sensitivities of Posterior Statistics

We focus on prior parameter dependence of inference as defined and used in the literature (Berger, Insua, and Ruggeri 2000; Giordano, Broderick, and Jordan 2017) in terms of the local derivatives (sensitivities) of a posterior statistic with respect to prior parameters

$$\frac{\partial \mathbb{E}_\pi[S(\boldsymbol{\theta})|Y, \boldsymbol{\eta}_0]}{\partial \boldsymbol{\eta}_0}. \quad (3)$$

Note that both $\boldsymbol{\eta}_0$ and $\mathbb{E}_\pi[S(\boldsymbol{\theta})|Y, \boldsymbol{\eta}_0]$ are high dimensional vectors, which in turns implies that the resulting expression for (3) is a high-dimensional matrix of dimension $\dim(S) \times \dim(\boldsymbol{\eta})$. For example, the sensitivity of all the posterior means with respect to all the prior mean parameters for the regression coefficients in a linear model with Gaussian errors.

While conceptually appealing, the sensitivities are usually unavailable in the closed-form given most models' complex stochastic dependence structure. Computational assessment faces additional problems due to the dimensionality of the problem in terms of outputs (set of posterior statistics) and inputs (set of prior parameters) in combination with the computational intensity of the original MCMC estimation. This poses challenges for existing methods such as finite differencing, symbolic

differentiation, and likelihood ratio methods to evaluate a broad set of prior parameters and a wide range of posterior output statistics.³ As a result, a common practice has been a less formal finite-difference approach based on rerunning the model under a small set of different prior input specifications (Chib and Ergashev 2009; Roos et al. 2015).

On the other hand, the proposed IPA approach uses the concept of perturbation analysis, that is, monitoring the path of the simulation as it evolves to estimate impacts of input parameter changes. From an algorithmic point of view relevant for such a perturbation analysis, an MCMC algorithm is simply a function or mapping that takes in a model of $\boldsymbol{\theta}$, a vector of hyper-parameters $\boldsymbol{\eta}_0$, and a starting vector $\boldsymbol{\theta}^0$ to obtain estimates of a range of statistics, based on the simulated posterior distribution from an MCMC algorithm. We show in Section 2.3 that by applying IPA we can express the local derivative (3) in terms of derivatives of the simulation draws as shown in Section 2.3.

The approach requires computing derivatives of simulated output. By definition, simulation is a way of evaluating integration. Derivate estimation via simulation inevitably requires an interchange of the order of differentiation and integration. In the context of IPA, we show in Section 2.4 that it is possible to obtain unbiased and consistent estimates of (3) via the IPA based sensitivity analysis as $\frac{\partial \mathbb{E}_\pi[S(\boldsymbol{\theta})|Y, \boldsymbol{\eta}_0]}{\partial \boldsymbol{\eta}_0} = \mathbb{E}_\pi[\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\eta}_0}|Y, \boldsymbol{\eta}_0]$, that is, it is valid to interchange the order of differentiation and integration in this context.

2.3. IPA Derivatives in Dependent Sampling

IPA constitutes one form of perturbation analysis that examines the behavior of a simulation path as the size of perturbations applied to input parameters approaches zero. This method has already found broad applications in classical simulation studies (Giles and Glasserman 2006; Glasserman and Liu 2010). Here, we extend the scope of IPA to encompass dependent sampling using Gibbs sampling algorithms. In this context, the transition kernel (1) is established on the basis of analytically known conditional posterior distributions of model parameters. We initiate by defining the central IPA output of interest using the Jacobian matrix denoted as $J_{\boldsymbol{\theta}^g}(\boldsymbol{\eta}_0)$. This matrix encapsulates the complete set of derivatives of the g th parameter draw $\boldsymbol{\theta}^g$ concerning the hyperparameters $\boldsymbol{\eta}_0$. It's important to note that, parametrically, the collection of data also constitutes an input of the algorithm. However, for the sake of simplicity, we omit the consideration of the impact of data in the current analysis. Further exploration of data impact is left for future research, hence, we drop Y in the notations.

MCMC algorithms present intricate simulation paths. Nevertheless, they can be broken down into a series of mappings. These mappings are designed to manage the transition kernel $\pi(\cdot|\boldsymbol{\theta}^{g-1}, \boldsymbol{\eta}_0)$ for generating $\boldsymbol{\theta}^g$ samples, and a third mapping for computing the posterior statistics of interest:

³An exception are sensitivities of the posterior mean with respect to prior mean parameters as computed in Pérez, Martin, and Rufo (2006) and Müller (2012).

1. A type one mapping updates the state parameter \mathbf{s} of the transition kernel, that is $f^1 : \mathbb{S}^\eta \times \Omega \rightarrow \mathbb{S}^s$ such that $\mathbf{s}^g = f^1(\boldsymbol{\eta}_0, \boldsymbol{\theta}^{g-1})$, with $\mathbb{S}^s \subset \mathbb{R}^{\dim(\mathbf{s})}$.
2. A type two mapping draws the model parameters with the updated state parameters via $f^2 : \mathbb{S}^s \times (0, 1)^{\dim(\omega)} \rightarrow \Omega$ such that $\boldsymbol{\theta}^g = f^2(\mathbf{s}^g, \omega^g)$ given the state parameters \mathbf{s} and a sequence of independent $U(0, 1)$.⁴

For example, suppose the transition kernel is multivariate Gaussian. At each iteration, the mean and covariance are matrix-valued functions (updated via f^1) of the draw from the last iteration and prior hyperparameters. Then a draw from this particular multivariate Gaussian is obtained given the updated mean and covariance using f^2 . This structure is easily generalizable to a multiple block setting. In a standard multivariate linear regression model with Normal errors and Normal priors on regression coefficients and Inverse Wishart prior for the covariance matrix of the errors, we would have a multivariate Gaussian update of the regression coefficients and an Inverse Wishart update of the covariance matrix. Again, type one mappings update the parameters in the two components of the transition kernel and type two mappings are used to generate each set of parameter draws, iteratively.

Considering a Gibbs algorithm expressed in terms of the two mappings, we can apply the chain rule to differentiate the Gibbs chain that generates the model parameter draws to obtain the expression for the main IPA output defined in terms of the Jacobian, $J_{\boldsymbol{\theta}^g}(\boldsymbol{\eta}_0)$, we have

$$J_{\boldsymbol{\theta}^g}(\boldsymbol{\eta}_0) = J_{f^2}(\mathbf{s}^g)J_{f^1}(\boldsymbol{\eta}_0) + J_{f^2}(\mathbf{s}^g)J_{f^1}(\boldsymbol{\theta}^{g-1})J_{\boldsymbol{\theta}^{g-1}}(\boldsymbol{\eta}_0). \quad (4)$$

Hence, the IPA analysis implies that the Jacobian matrix of sensitivities of the parameter draws with respect to the prior parameters can be expressed in terms of the Jacobian of the second mapping with respect to the state parameters and Jacobians of the first mapping with respect to the state parameters, prior parameters, and previous draws. Note that the expression (4) contains a path-wise and a chain-rule components. The path-wise component arises from the *direct* influence of the hyperparameters on the transition kernel, that is, the random number generation algorithm is a function of the hyper-parameters. In addition, they also affect the $(g - 1)$ th draw, which determines the state parameters of the transition kernel at the current iteration g .

Finally, in order to compute posterior statistics, a type three mapping in Gibbs inference involves the computation of samples statistics $S(\boldsymbol{\theta}^g)$ given the draws $\boldsymbol{\theta}^g$. Bayesian inference typically involves the analysis of the posterior distribution, with particular attention given to the posterior mean. In this case, identity mapping is the third type of mapping in the algorithm. Hence, the sensitivity of posterior statistics w.r.t. the complete set of prior inputs can be computed by passing the derivative operator through S given the already available $J_{\boldsymbol{\theta}^g}$, and averaging over the sample.

We can now introduce the following measure of the local input sensitivity the posterior statistics considered in the literature based on IPA analysis (IPA derivatives), where sensitivities

of posterior statistics of the model parameters with respect to the prior hyperparameters are given by

$$J_{\hat{S}(\boldsymbol{\theta})}(\boldsymbol{\eta}_0) = \frac{\partial \mathbb{E}_{\hat{\pi}}[S(\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\eta}_0]}{\partial \boldsymbol{\eta}_0} = \frac{1}{G} \sum_{g=B+1}^{G+B} J_S(\boldsymbol{\theta}^g)J_{\boldsymbol{\theta}^g}(\boldsymbol{\eta}_0) \quad (5)$$

where $\mathbb{E}_{\hat{\pi}}$ reflects that the expectation of S is obtained from the empirical distribution formed with posterior draws, and G is the total number of iterations after the burn-in period B .

2.4. Convergence and Asymptotics

In the previous section, we have introduced the idea of IPA to Bayesian MCMC-Gibbs simulation settings given the input parameters $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}^0$. We are interested in the derivatives of some posterior statistics $\alpha(\boldsymbol{\eta}_0, \boldsymbol{\theta}^0)$ written as expectations over some real-valued objective functions. When a sample is drawn independently, as in the classical simulation setting, convergence properties of IPA derivatives in the form of equation (5) are implied by the basic Law of Large Numbers. In this section, we extend convergence properties to dependent sampling via Gibbs MCMC settings, where the key focus is to estimate the derivative of α as defined in (2) with the expectation taken over the true posterior distribution. Note that the assumptions we make in the following section are regarding the Markov transition kernel, that it naturally flows on to the impact of $\boldsymbol{\theta}^0$.

2.4.1. Asymptotic Unbiasedness

Sampling from the joint posterior distribution using MCMC techniques starts after discarding the first B draws from the algorithm. That is because generally we assume that from $B + 1$ onwards, we are sampling from the exact posterior distribution. In fact for an ergodic Markov chain, as $B \rightarrow \infty$, we reach the stationary distribution, and $\boldsymbol{\theta}^{B+1}$ becomes unbiased, that is, $\lim_{B \rightarrow \infty} \mathbb{E}[\boldsymbol{\theta}^{B+1}] = \mathbb{E}[\boldsymbol{\theta}|Y]$. Hence, it is critical for analysis to collect sensitivity estimates that is unbiased from $B + 1$ onwards. The basic idea behind the Gibbs sampler, is that a posterior statistics α can be viewed as $\alpha(\boldsymbol{\eta}_0) = \int_{\Omega} S(\boldsymbol{\theta})d\pi_{\boldsymbol{\eta}_0}(\boldsymbol{\theta})$ which is the expectation of an integrand S of the sample path, $\boldsymbol{\theta}$, with respect to a posterior probability measure $\pi_{\boldsymbol{\eta}_0}$ over some measurable space (Ω, \mathcal{F}) .⁵ In general, one cannot differentiate the integrand in Gibbs settings, as $\pi_{\boldsymbol{\eta}_0}$ depends on $\boldsymbol{\eta}_0$. Note that to simplify notation we are suppressing the conditioning on the data.

The choice of sample space Ω and its corresponding sample path is crucial in IPA (L'Ecuyer 1990). In fact, for all simulation settings, all random variables are generated from $\omega \sim U(0, 1)$ variates. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\pi})$ denote a probability space in which the random element is a sequence of independent $U(0, 1)$ variates. Because Gibbs samples are drawn dependently via a transition kernel, $\pi(\cdot|\cdot, \boldsymbol{\eta}_0)$, that is $\boldsymbol{\theta}^{g+1}$ is drawn from $\pi(\cdot|\boldsymbol{\theta}^g, \boldsymbol{\eta}_0)$, we shall use $(\Omega^g, \mathcal{F}^g, \mathbb{P}_{\boldsymbol{\eta}_0}^g)$ to denote the conditional probability triplet.

Assumption 1. For $(\Omega^g, \mathcal{F}^g, \mathbb{P}_{\boldsymbol{\eta}_0}^g)$, assume that

- (i) the measurable space $(\Omega^g, \mathcal{F}^g) = (\Omega, \mathcal{F})$

⁴In practice, we do not have to work in this microscopic fashion, as long as we can find a representation of the sample path that its probability the measure does not depend on the prior hyperparameters.

⁵Note here we dropped \mathbf{y} in the input vector, as it is not the main focus of this article.

- (ii) for every measurable set \mathcal{A} of (Ω, \mathcal{F}) , the transformation $\phi_{\eta_0, \theta} = f^2 \circ f^1$ is such that

$$\mathbb{P}_{\eta_0}^g(\mathcal{A}) = \tilde{\pi}(\phi_{\eta_0, \theta}^{-1}(\mathcal{A})).$$

To apply IPA, the standard regularity conditions are needed for the algorithm. Since the algorithm is constructed via a stochastic representation of the transition kernel in Gibbs, we have the following conditions similar to those in the classical simulation setting.

Assumption 2. The following statements are considered under $(\Omega^g, \mathcal{F}^g, \mathbb{P}_{\eta_0}^g)$.

- (i) \mathbb{S}^η is compact.
- (ii) For all $\eta_0 \in \mathbb{S}^\eta$ and $\theta \in \Omega$, the transformation $\phi_{\eta_0, \theta}$ is differentiable in both η_0 and θ with probability one. For all $\eta_0, \eta_0^* \in \mathbb{S}^\eta$ and $\theta, \theta^* \in \Omega$, there exist random variables C_1 and C_2 such

$$\left\| \phi_{\eta_0, \theta}(\omega) - \phi_{\eta_0^*, \theta}(\omega) \right\| \leq C_1 \left\| \eta_0 - \eta_0^* \right\|$$

$$\left\| \phi_{\eta_0, \theta}(\omega) - \phi_{\eta_0, \theta^*}(\omega) \right\| \leq C_2 \left\| \theta - \theta^* \right\|$$

with $\mathbb{E}[C_1] < \infty$, $\mathbb{E}[C_2] < 1$.⁶

- (iii) Let $D_S \subseteq \Omega$ denote the compact set at which S is differentiable, $\Pr(S(\theta) \in D_S) = 1$ and S is Lipschitz for all $\theta \in \Omega$.

Compactness is a standard assumption here for sensitivity analysis and estimation. Assumptions (ii) and (iii) are concerned with the smoothness of the algorithm and its capacity to allow the passing through of derivative operators. Assumption (ii) is also related to convergence and the boundedness of the derivative as $B \rightarrow \infty$ and ensures that $\frac{\partial \theta^B}{\partial \theta^j} = \prod_{i=j}^{B-1} \frac{\partial \theta^{i+1}}{\partial \theta^i}$ pre-multiplied with $\frac{\partial \theta^{j+1}}{\partial \theta^j}$, is essentially a product of matrix valued random variables (see Raghunathan (1979) for a theoretical discussion). Note this is not a very strong condition since under a geometric convergent chain these partial derivative ought to converge to zero as $B \rightarrow \infty$ at a geometric rate almost surely.

Theorem 1 (Asymptotic Unbiased IPA Derivatives). Under Assumptions (1) and (2), and for all $\eta \in \mathbb{S}^\eta$, $\left\| \pi^B(\cdot | \theta^0, \eta_0) - \pi(\cdot | \eta_0) \right\| \rightarrow 0$ we have unbiasedness for the limit of equation (5) as $B \rightarrow \infty$ of $\frac{\partial}{\partial \eta_0} \alpha(\eta_0)$, that is, $\mathbb{E}[\lim_{B \rightarrow \infty} J_{G,B}(\eta_0)] = \frac{\partial}{\partial \eta_0} \alpha(\eta_0)$.

See the proof in Appendix A.

2.4.2. Asymptotic Properties as $G \rightarrow \infty$ and $B \rightarrow \infty$

We have shown the estimator to be unbiased as $B \rightarrow \infty$. In this section, we discuss its property as $G \rightarrow \infty$. In the classical simulation setting, the consistency and asymptotic normality are straightforward to show for the IPA derivatives, given its unbiasedness and diminishing finite variance. In the Gibbs context, however, dependent samples are drawn, and their derivatives

are calculated along the Markov chain. Hence, the variance of the sample mean also includes covariance terms. Therefore, our proof of the following limiting theorem relies on the construction of Mixingale type stochastic processes. Their limiting behavior is discussed in detail in the time series literature (see De Jong 1995).

Theorem 2 (Consistency for the IPA Derivatives). Under assumption (1) and (2), we have $J_{G,B}(\eta_0) \xrightarrow{P} \frac{\partial}{\partial \eta_0} \alpha(\eta_0)$ for all $\eta_0 \in \mathbb{S}^\eta$ as $G \rightarrow \infty$ and $B \rightarrow \infty$.

The proof is provided in Appendix A.

The established asymptotic properties of IPA-based derivatives hold true across algorithms applied to diverse models. We will showcase the effectiveness of this method by using it within linear regression frameworks, both in real-world applications and simulation studies. Furthermore, this method can be extended to a broad array of state space models, encompassing models like the time-varying coefficient Vector Autoregression (VAR) (Primiceri 2005) and dynamic factor models (Diebold and Li 2006), each with varying prior specifications. Notably, the standard algorithm for these model types, typically founded on the Kalman Filter (Carter and Kohn 1994), adheres to the necessary regularity conditions, thereby making the IPA approach applicable.

An inherent limitation of the IPA approach lies in its requirement for algorithm smoothness. Algorithms of the rejection type, including the widely used Metropolis-Hastings (MH) algorithm, generate nonsmooth simulation paths. This conflicts with the Lipschitz assumptions established in our theoretical framework. Nevertheless, in situations involving low dimensions and assuming the density itself is continuously differentiable, we can resort to numerical integration techniques to derive distributional derivatives. In Appendix D, we present an illustration of a possible expansion of the standard IPA setup to MCMC settings where the full conditional distribution remains unknown for a specific model parameter. We have also implemented an IPA based approach within Metropolis-within-Gibbs and Slice-within-Gibbs algorithms and demonstrated the performance in a simulation study (see Appendix B). Note that algorithms such as unadjusted Hamiltonian Monte Carlo (Bou-Rabee, Eberle, and Zimmer 2020) and Langevin sampling (Durmus and Moulines 2017) fulfill the required smoothness assumption.

2.5. Implementation

To address the reliance of IPA on the exact evaluation of complex derivative operations on the simulation algorithm, the automatic differentiation (AD) approach was introduced in classical simulation community by Griewank and Walther (2008), now a standard computational tool for computing derivatives of smooth functions (Homescu 2011).⁷ While AD has been included as part of common statistical libraries, such as PYMANOPT for optimization of manifolds implemented in Python (Townsend,

⁶Here, we consider the euclidean norm for the vectors.

⁷Since the derivative estimate computed using AD is identical to the ordinary pathwise estimates, its potential advantage is, therefore, computational rather than statistical.

Koep, and Weichwald 2016) and Stan for approximate inference (Giordano, Broderick, and Jordan 2017), these are not designed to handle some of the specific computational challenges of Gibbs sampling. Since it fails to exploit substantial gains in computational efficiency from matrix operations, a naive application of such standard element by element AD implementations for a IPA-based sensitivity analysis in Gibbs settings can be computationally prohibitively expensive. Our implementation, therefore, exploits a Vector-based AD approach discussed in Kwok, Zhu, and Jacobi (2022) to facilitate the efficient computation of the complete set of prior parameter sensitivities of the MCMC draws alongside the Gibbs sampling iterations. We refer interested readers to the R package (Kwok, Zhu, and Jacobi 2020) for an implementation of Vectorized AD in a general multivariate regression context. The code is also available as part of the supplementary materials.

3. Simulation Study

We illustrate the performance and use of the introduced IPA approach for prior parameter sensitivities of posterior statistics from Gibbs inference in a multivariate regression framework. Our experimental design is based on a standard BVAR model and explores prior parameter sensitivities in the context the common use of shrinkage priors to address overparameterization arising from the moderate size of many macroeconomic datasets (Karlsson 2013). Within this setting we first benchmark IPA derivative estimates against two alternative computational methods to demonstrate the unbiasedness and high precision of IPA estimates implemented. This part of the simulation focuses on sensitivities of posterior mean and variance statistics to aid benchmarking given limitations of existing methods. In the second part of the simulation study we exploit the flexibility and efficiency of the proposed IPA tool by contrasting patterns of prior parameter dependence under shrinkage with more uninformative prior parameter settings, and by expanding the set of considered sensitivities to assess sensitivity patterns of key features in VAR-based policy analysis, such as impulse responses to policy shocks.

3.1. Experiment Design

Consider the VAR(q) model, which can be expressed as

$$\mathbf{y}_t = \mathbf{b}_0 + \sum_{l=1}^q \mathbf{y}_{t-l} \mathbf{B}_l + \boldsymbol{\epsilon}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t ; \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad (6)$$

where $t = 1, 2, \dots, T$, $\mathbf{X}_t = \mathbb{I}_n \otimes (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-q})$ and $\boldsymbol{\beta} = \text{vec}([\mathbf{b}_0, \mathbf{B}_1, \dots, \mathbf{B}_q]')$. The row-stacked vector of intercepts and VAR (lag) coefficients is denoted as $\boldsymbol{\beta} \in \mathbb{R}^{k_\beta}$, where $k_\beta = n(q+1)$. The latter part of VAR (lag) coefficients captures the dynamic interdependence among the time series. We revisit the BVAR setting in the empirical analysis. For the simulation study, we consider two data generating processes:

- DGP1: Independent normal VAR(2) with $\mathbf{b}_0 = \mathbf{1}_n$, $\mathbf{B}_l = \frac{1}{l+1} \mathbb{I}_n$ and $\boldsymbol{\Sigma} = 0.01 \mathbb{I}_n$;
- DGP2: General normal VAR(2) with $\mathbf{b}_0 = 0.01 \times \mathbf{1}_n$, \mathbf{B}_l and $\boldsymbol{\Sigma}$. The diagonal elements of the first VAR coefficient

\mathbf{B}_1 are iid uniform $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are $\mathcal{U}(-0.2, 0.2)$. All elements of the higher VAR coefficient matrix \mathbf{B}_l are iid $\mathcal{N}(0, 0.05^2/l^2)$, where l is the lag length. The error covariance matrix $\boldsymbol{\Sigma}$ is generated from the Inverse-Wishart distribution $\mathcal{IW}(n+10, 0.07 \mathbb{I}_n + 0.03 \mathbf{1}_n \mathbf{1}_n')$.

Under both data designs the number of endogenous variables n is set at 3 and the time length T is set at 600. For the simulation study 100 samples are generated under each DGP with the initial value of \mathbf{y}_0 set as the stationary mean of the corresponding process.

For each dataset simulated from the true DGP, we will fit a VAR(p) model, with the possible options for the lag order p of fitted model being 1, 2, or 4. We proceed under standard independent (conditional conjugate) prior assumptions with a Normal prior on $\boldsymbol{\beta}$, $N(\boldsymbol{\beta}_0, \mathbf{B}_0)$, and an Inverse Wishart prior on $\boldsymbol{\Sigma}$, $IW(\nu_0, \mathbf{S}_0)$, with the exact implementation given in Appendix B.⁸ Within this framework, we set the prior mean $\boldsymbol{\beta}_0$ to be zeros and proceed under different prior parameter specifications for the prior variances of the lag coefficients (\mathbf{B}_0), encompassing both uninformative and shrinkage settings. For the uninformative prior, we set $\mathbf{B}_0 = 100 \mathbb{I}_{k_\beta}$ ($k'_\beta = n(np+1)$), a standard choice in many practical applications. Our main prior specification is a Minnesota-type shrinkage prior that shrinks higher lag-coefficients to zero (Koop and Korobilis 2010; Karlsson 2013), thus, favoring more parsimonious models and avoiding overfitting the data (Morley and Wong 2017). In particular, we set \mathbf{B}_0 as a diagonal matrix with different formulations for lag and intercept coefficient variances such that $\mathbf{B}_{0,ii} = \kappa_0/l^2$ if i corresponds to the own l th lag, $\kappa_1 s_i^2 / (l^2 s_j^2)$ if i corresponds to l th lag coefficients of j variable in the i th equation and $\kappa_2 s_i^2$ if it is intercept. We set $\mathbf{S}_0 = \kappa_3 \mathbb{I}_n$, with $\kappa_0 = 0.1^2$, $\kappa_1 = 0.1^2$, $\kappa_2 = 100$, and $\kappa_3 = 1$.

3.2. Benchmarking IPA Derivatives

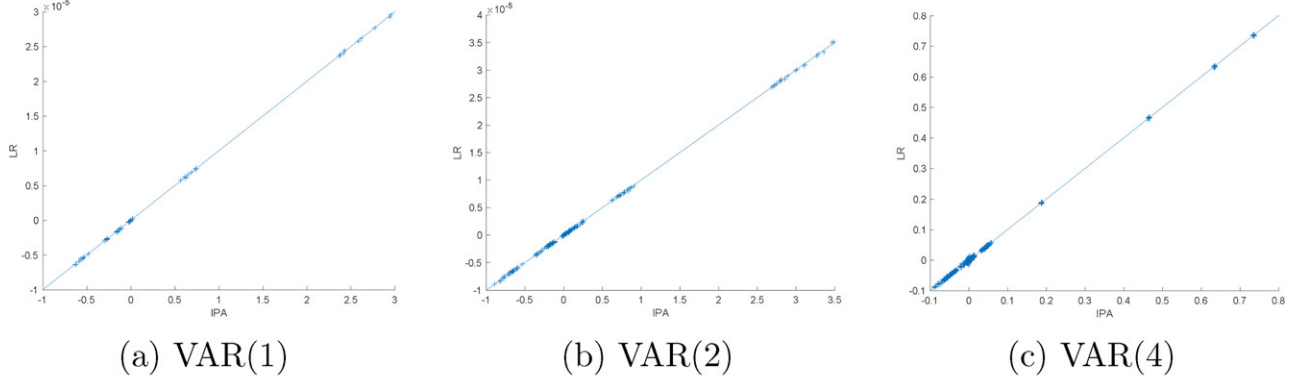
Following the classical simulation literature, we compare the IPA derivative estimates against the two popular alternative numerical derivative estimation methods, the likelihood ratio (LR) and the numerical finite-difference (FD) methods (Glasserman 2013). Like IPA, the LR method is unbiased, but relies on analytically evaluating and computing derivatives with respect to the loglikelihood.⁹ Within the MCMC context its implementations so far are restricted to the derivative of the posterior mean with respect to the prior mean within the exponential family of distributions (Pérez, Martin, and Rufo 2006; Müller 2012). Appendix B provides more details on the LR and FD implementations.

All presented benchmarking results refer to VAR inference under the Minnesota shrinkage prior for data simulated under

⁸Compared to the more restrictive natural conjugate joint Normal-Wishart prior formulations, favored in large Bayesian VAR analysis to circumvent computational challenges of Gibbs inference in such settings, it allows for asymmetry in the prior and posterior (Carriero, Clark, and Marcellino 2019).

⁹We acknowledge the fact that there is a recent generalization of the LR method within the classical simulation context (Peng et al. 2018) which has not been applied in the MCMC setting.

Comparing Derivative Estimates: IPA and LR



Comparing Precision (SD) of Derivative Estimates: IPA and LR

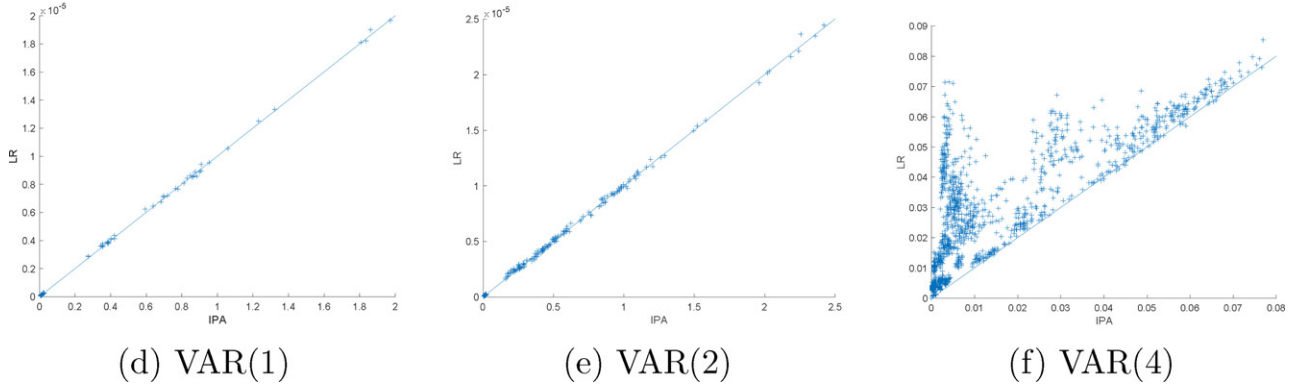


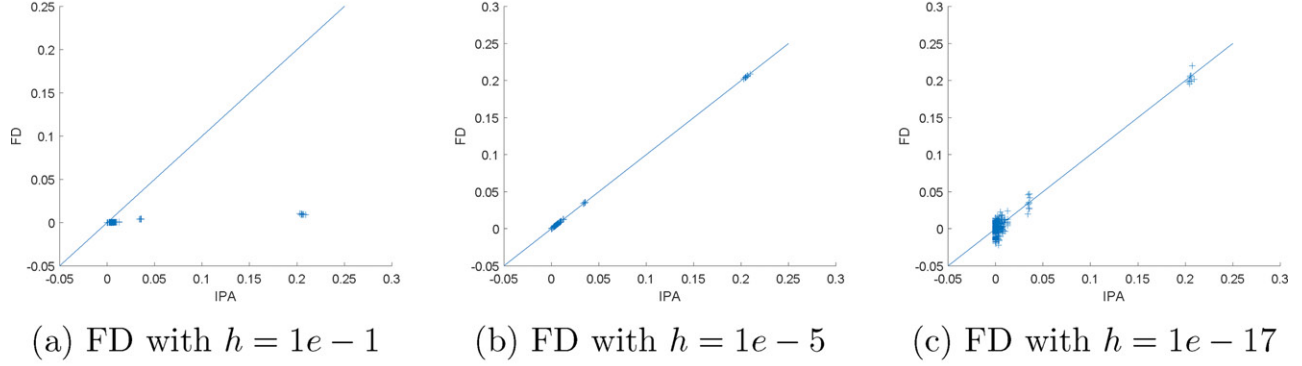
Figure 1. Comparing estimates of posterior mean derivatives (top panel) and their corresponding precision in terms of estimator standard deviations (bottom panel) obtained through the IPA on the x-axis and the LR on the y-axis. The findings are derived from 100 samples originating from DGP 2, using VAR(p) ($p = 1, 2, 4$) fitting under Minnesota shrinkage priors.

DGP 2. We begin with the comparison of the derivative estimates for $\frac{\partial}{\partial \beta_0} \mathbb{E}[\beta | y, \eta_0]$ via the IPA and LR methods. Figure 1 compares the posterior mean derivatives estimates under AD and LR in the top panels (a)–(c) and precision of the corresponding precisions of the derivative estimates in terms of their standard deviations in the bottom panel (d)–(e). First, both IPA and LR produce unbiased derivative estimates, with the points reflecting the derivative estimates under both methods closely tracing the 45-degree line. Second, the bottom set of graphs shows that the IPA estimates are estimated more precisely for VARs with more than one lag with most points well above the 45-degree line reflecting a higher standard deviation of the LR estimates. This pattern is in accordance with the large evidence from the classical simulation literature that shows that the LR method generally produces derivative estimates with larger standard deviations than that of the IPA estimator (Cui et al. 2020).

Next, we turn to the derivatives of the posterior variances $\frac{\partial}{\partial \text{diag}(\mathbf{B}_0)} \text{var}(\beta | y, \eta_0)$ that can be benchmarked against the more general FD methods. Estimates of the fully numerical FD approach are biased and depend on the chosen bump size in the implementation. Hence, FD provides an approximation subject to the widely documented variance-bias tradeoff, also evident in our simulation results where we benchmark the IPA estimates with estimates obtained under the FD implementation under various bump sizes. Note, that since FD assesses derivatives

with respect to one prior input parameter at a time, it requires many reruns of MCMC under some chosen bump size with the fixed random number generators. While different bump sizes could be chosen for different dimensions, the complexity of the problem makes such choices difficult in practice (Glasserman 2013). Figure 2 compares IPA derivatives with FD derivatives under three bump sizes (1e-1, 1e-5, and 1e-17). The variation in the patterns across the columns reflects the dependence of the FD performance, in terms of both point estimates (graphs (a)–(c)) and their precisions (graphs (d)–(e)), on the chosen bump size and the inherent variance-bias tradeoff. If the bump size is too small, then the FD point estimates of the sensitivities are reasonable (c), although with considerably more variation around the 45-degree line than LR estimates, and suffering from a low precision (f) due to the large variance and floating point error. On the other hand, if the bump size is larger, we observe a substantial bias in the FD estimates (2a) combined with a very high precision (2d). There is always a variance/bias tradeoff in the application of FD with smaller bump values leading to smaller bias but large variances, and vice versa for larger bump values. In practice, the appropriate choice is hence difficult to determine. In our particular setting, a moderate bump size of (1e-5) appears close to the optimal choice as it results in FD very close to AD. Note that we can only conclude this in hindsight after comparison with the unbiased IPA estimates and after some fine-tuning by trying many bump sizes.

Comparing Derivative Estimates: IPA and FD



Comparing Precision (SD) of Derivative Estimates: IPA and FD

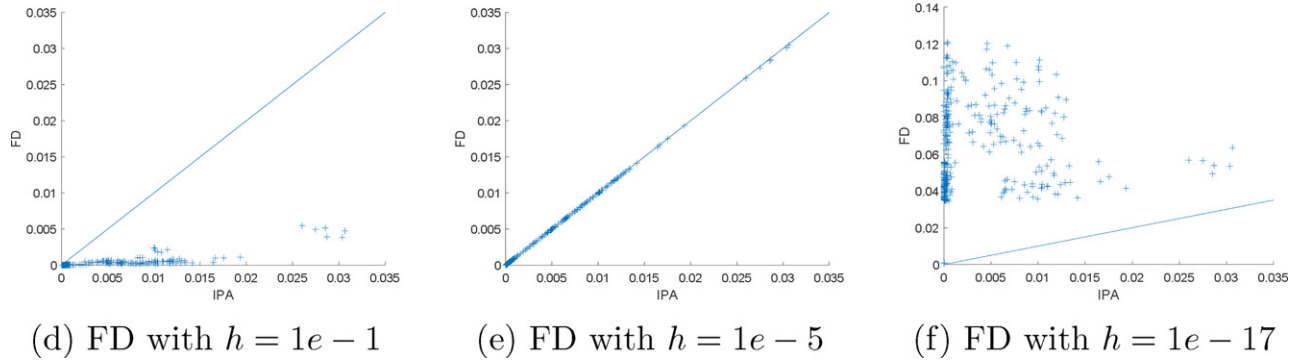


Figure 2. Comparing estimates of posterior variance derivatives and their corresponding precision from IPA and FD under various bump sizes ($1e-1$, $1e-5$, and $1e-17$). Results are from fitting of a VAR(2) model under Minnesota shrinkage priors, using 100 samples drawn from DGP2.

Overall, the benchmark results demonstrate that the proposed IPA approach yields estimates consistent with unbiased LR estimates with stabler results, in line with findings from a classical simulation context (Glasserman 2013). In terms of computational time, IPA takes less than a second for one MCMC parameter inference and gradient computation, while FD takes roughly nine seconds on average. This computational gain is achieved in our context using AD's advances, which execute all the derivative computations alongside the parameter inference from one MCMC chain. Due to the relatively simple structure of the posterior mean with respect to the prior mean, derivative estimation under the LR method only requires the prior and posterior covariance matrices and takes less than half a second. In more complicated settings, particularly those involving intractable likelihoods, the LR approach requires much more effort as it relies on taking derivatives of the density function and generally does not lead to a simple analytical expression as in the presented example.

3.3. Sensitivity Analysis with IPA Derivates

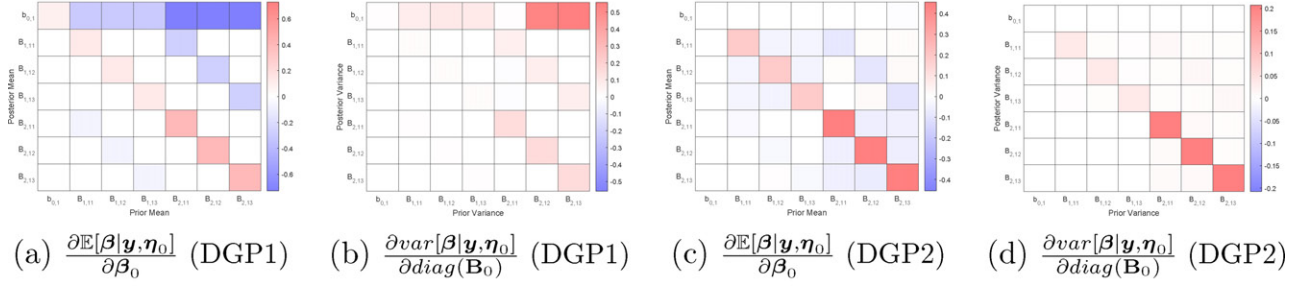
We now turn to exploiting the generality and computational efficiency of the proposed IPA approach to present patterns in prior parameter dependence for an expanded set of posterior measures, beyond lag coefficients, that relate directly to relevant features in a VAR-based policy analysis. However, we still begin by presenting the estimates for the IPA sensitivities of the regression coefficients also presented in the benchmarking exercise.

Figure 3 depicts heatmaps illustrating the sensitivities of posterior means and variances of the VAR coefficient matrix β within the context of inference using VAR(2) and VAR(4) models. These models incorporate the Minnesota prior and are applied to datasets sourced from DGP1 and DGP2. For clarity of exposition, we only present the sensitivities of posterior means and variances of VAR coefficients in the first equation with respect to the corresponding prior means and variances. Please refer to Appendix E for the heatmap in its entirety. Upon examining both rows of Figure 3, it becomes evident that common sensitivity patterns emerge across the VAR(2) and VAR(4). The results clearly indicate that as the model increases in size, that is, the number of VAR coefficients changes from 21 to 39 as transitioning from VAR(2) to VAR(4), the influence of the prior parameters becomes more pronounced while preserving a similar pattern.

Figure 4 showcases the corresponding sensitivities when conducting inference under an uninformative prior (For more details, please refer to Appendix E). Notably, these sensitivities are markedly reduced compared to the outcomes of the previous analysis, thereby confirming the substantial impact of prior shrinkage on posterior statistics.

Next, we present IPA derivative estimates for IRFs and FEVDs. To compute the IRF and FEVD, we focus on the Cholesky decomposition of the structural VAR. This allows us to demonstrate that IPA serves as a tool for examining the influence of prior parameters on policy inference through IRF derived from orthogonal shocks, as implied by the Cholesky

Sensitivities of Regression Coefficients: Minnesota Prior (VAR(2))



Sensitivities of Regression Coefficients: Minnesota Prior (VAR(4))

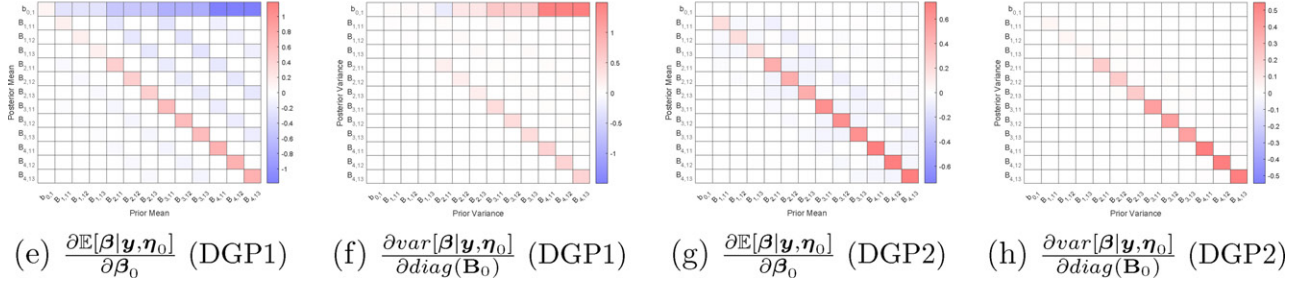
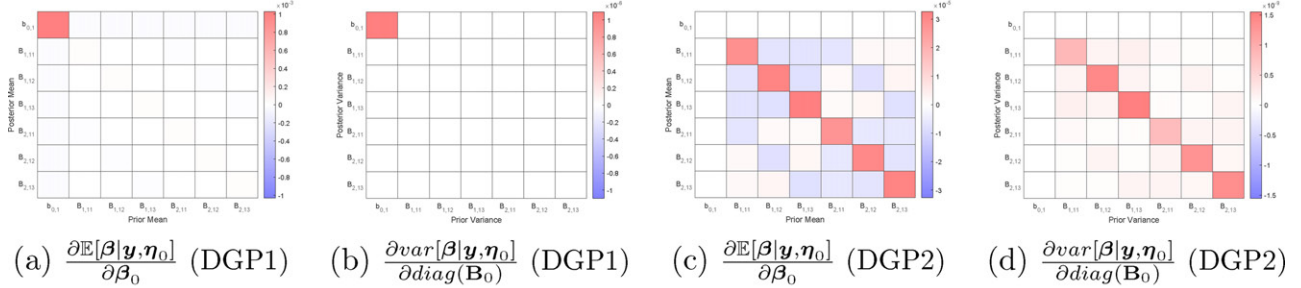


Figure 3. IPA estimates of posterior mean and variance derivatives under Minnesota shrinkage priors for data simulated under DGP1 and DGP2. Presented estimates are for a VAR(2) or VAR(4) model for the sensitivities of posterior means and variances of VAR coefficients in the first equation with respect to the corresponding prior means and variances.

Sensitivities of Regression Coefficients: Uninformative Prior (VAR(2))



Sensitivities of Regression Coefficients: Uninformative Prior (VAR(4))

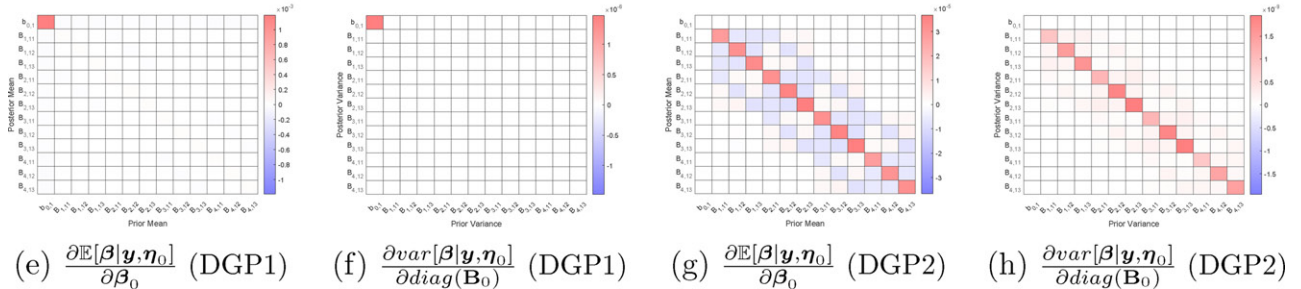


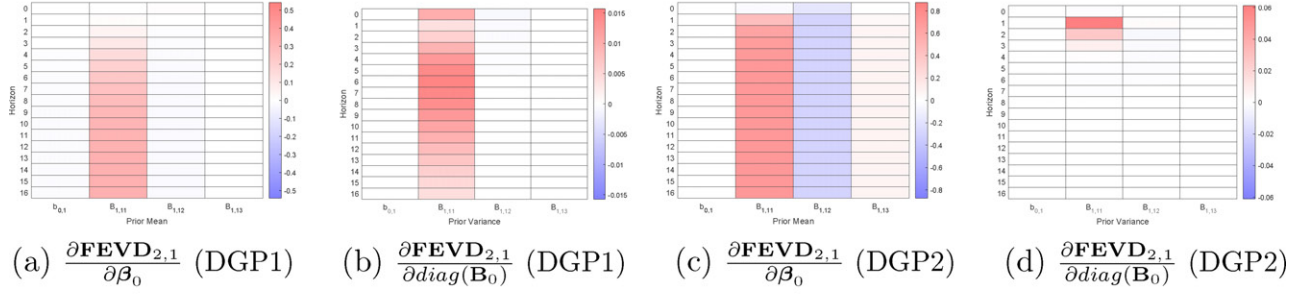
Figure 4. IPA estimates of posterior mean and variance derivatives under the Uninformative prior for data simulated under DGP1 and DGP2. Presented estimates are sensitivities of posterior means and variances of VAR coefficients in the first equation with respect to the corresponding prior means and variances for inference under a VAR(2) model.

decomposition. Note that the algorithm for computing the Cholesky decomposition is inherently smooth, that is, satisfying the regularity conditions of our theoretical results.

Let $\text{IRF}_{i,j,t}(\beta, \Sigma)$ denote the algorithm for calculating the IRF of the i th variable concerning the shock to the j th variable at time t , where $\{\beta, \Sigma\} \sim \pi(\cdot | \text{data}, \eta_0)$. We introduce

the prior sensitivities in terms of the mean IPA derivatives with respect to the relevant set of prior parameters η_0 , that is, $\frac{\partial}{\partial \eta_0} \mathbb{E}_\pi [\text{IRF}_{i,j,t}(\beta, \Sigma)]$, obtained by averaging samples $\frac{1}{S} \sum_{g=B+1}^{S+B} \frac{\partial}{\partial \eta_0} \text{IRF}_{i,j,t}(\beta^g, \Sigma^g)$. We derive the sensitivities of FEVD in a similar manner. The interchange of the integration

IPA Derivatives of Forecast Error Variance Decomposition (VAR(1))



IPA Derivatives of Forecast Error Variance Decomposition (VAR(2))

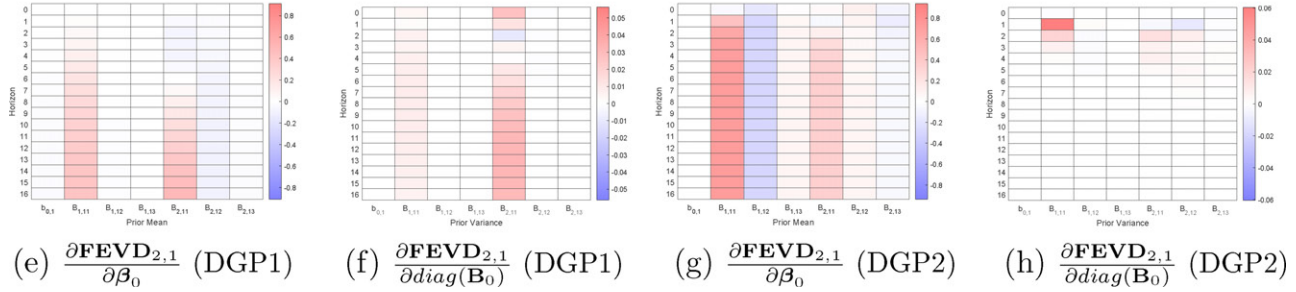


Figure 5. IPA estimates of posterior forecast error variance decomposition $\text{FEVD}_{2,1,t}$ ($t = 0, 1, \dots, 16$) with respect to both prior means and prior variances of VAR coefficients in the first equation under Minnesota shrinkage prior for data simulated from DGP1 and DGP2.

(expectation) and the derivative computation arises as a consequence of the smoothness in the Gibbs transition kernel and the computation of IRFs/FEVDs.

For ease of understanding, Figure 5 displays sensitivity estimates of $\text{FEVD}_{2,1,t}$ for $t = 0, 1, \dots, 16$ along the y -axis with respect to priors of VAR parameters in the first equation, representing the FEVD of the second variable under the influence of the first shock. We further present an entire heatmap illustrating sensitivity estimates of $\text{FEVD}_{2,1,t}$ for $t = 0, 1, \dots, 24$ with respect to prior means and variances of all VAR coefficients β in Appendix E. As the FEVD is independent of the intrinsic magnitudes of the DGPs, direct comparisons between DGP 1 and DGP 2 results are feasible. In the context of DGP 2, when compared to DGP 1, $\text{FEVD}_{2,1,t}$ demonstrates heightened sensitivities across a wider array of prior parameters. This is likely attributed to the interdependencies among variables inherent to DGP 2.

Figure 6 showcases sensitivity estimates for the IRF under the same conditions as presented in Figure 5 (For full details, please refer to Appendix E). The lower panels (e)–(h) present “relative” IPA estimates, wherein each sensitivity estimate is scaled by a factor corresponding to the value of $\text{IRF}_{2,1,t}$ itself.¹⁰ These relative IPA estimates illustrate that while the IPA derivatives themselves may appear relatively small, they hold significant importance when considered in relation to the magnitudes of the original posterior statistics.

¹⁰More generally, a relative IPA derivative is defined as the ratio of the sensitivity and the corresponding IRF, that is,

$$\frac{\partial}{\partial \eta_0} \mathbb{E}_\pi [\text{IRF}_{ij,t}(\beta, \Sigma)] / \mathbb{E}_\pi [\text{IRF}_{ij,t}(\beta, \Sigma)].$$

4. Application

In this section, we employ the IPA-based derivative approach to investigate the sensitivity of the estimated IRFs to prior parameter choices in the context of an orthogonal government spending shock. We focus on a fiscal policy Bayesian VAR (BVAR) model applied to U.S. data, a methodology akin to the frameworks used in Jarociński and Marcet (2019) and Ramey (2019). While VARs and their associated IRFs have a longstanding history in macroeconomic policy analysis (Blanchard and Perotti 2002), they have garnered renewed attention among researchers due to the heightened employment of fiscal stimulus in response to the Covid pandemic. Bayesian VAR analysis remains a favored approach for policy evaluation and forecasting, offering a variety of prior specification options within the literature. However, the incorporation of new data points from the Covid era introduces sparks renewed robustness concerns both to modeling and prior choices (Hauzenberger et al. 2021).

Establishing the suitability of prior assumptions on the autoregressive parameters in a specific application remains a challenge with existing tools (Amir-Ahmadi, Matthes, and Wang 2018) and restricts the assessment of impacts of such prior parameter choices on IRF inference. Recent work by Jarociński and Marcet (2019) compares IRF estimates by repeated estimation under alternative sets of priors and finds that magnitude and dynamics of output responses from an orthogonal spending shock can vary considerably under different prior assumptions on the mean coefficients. The IPA-based derivative analysis introduces the option of a comprehensive and precise local sensitivity analysis, by identifying both key prior parameters driving prior sensitivity and the overall local sensitivity. Our analysis focuses on the impacts of prior mean and variance parameter changes to implied GDP responses from an orthogonal fiscal

IPA Derivatives of Impulse Responses (VAR(2))

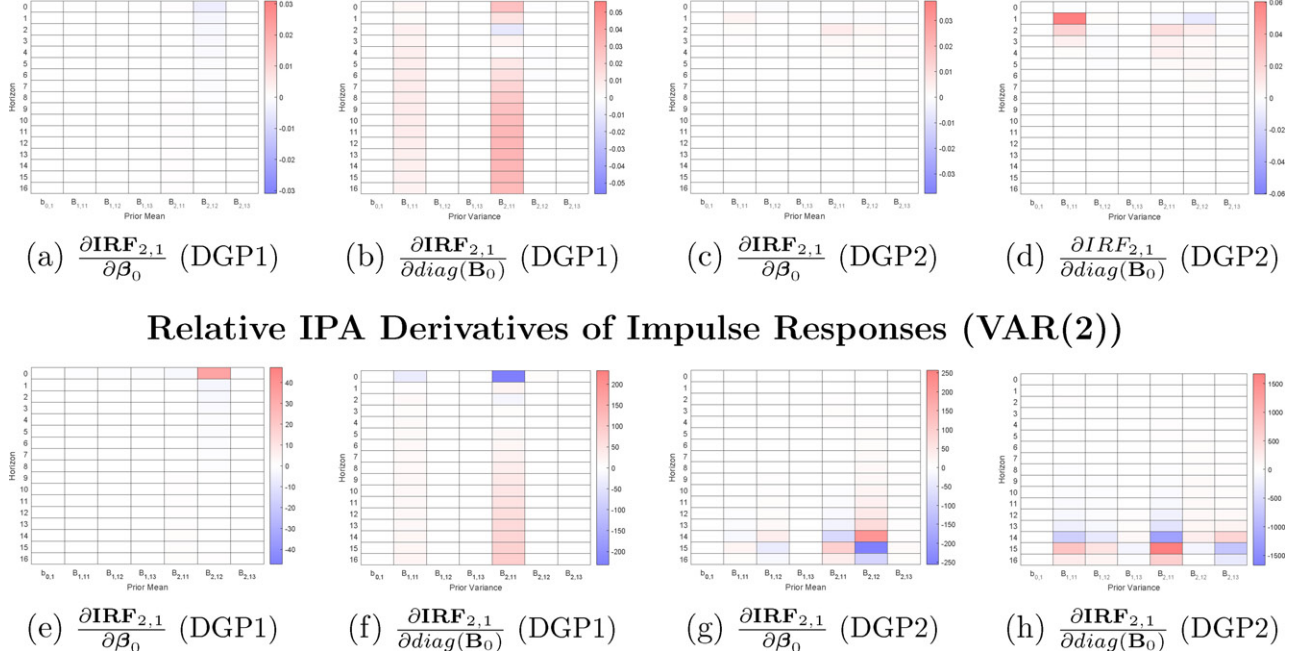


Figure 6. IPA and “Relative” IPA estimates of posterior $\text{IRF}_{2,1,t}$ ($t = 0, 1, \dots, 16$) with respect to prior means and prior variances of VAR coefficients in the first equation under Minnesota shrinkage prior for data simulated from DGP1 and DGP2.

policy shock that are estimated from a BVAR under a common Minnesota shrinkage prior.

4.1. Policy BVAR with Minnesota Prior

We consider a basic fiscal policy VAR for Government spending, GDP and Federal tax revenue based on a subset of the data used by Ramey (2019). The variables are defined in terms of the nominal government purchases, the natural logarithm of real GDP over the total population and tax revenues for the period from 1939 to 2015.¹¹ The analysis is based on a 3-variable 4-lag Bayesian VAR as specified in (6). A similar model has also been considered in Jarociński and Marcet (2019).

We proceed under an independent Normal-Wishart prior set-up also considered in the simulation exercise. For the Normal prior on the regression coefficients β we specify a common Minnesota shrinkage prior (Koop and Korobilis 2010; Karlsson 2013) to avoid overfitting the data by setting prior means at zero. The prior covariance matrix \mathbf{B}_0 is set as a diagonal matrix with a specific structure to shrink coefficients related to higher lag orders stronger to zero. In particular, we consider $\mathbf{B}_{0,ii} = \kappa_3 \frac{s_i^2}{p^2 s_j^2}$ for the l th lag coefficient of variable j in the i th equation, $\mathbf{B}_{0,ii} = \kappa_1 \frac{1}{p^2}$ for the own lag variables, κ_2 if it is the intercept, and $S_0 = \kappa_0 \mathbb{I}_n$. We set $\kappa_0 = 1$, $\kappa_1 = 0.2^2$, $\kappa_2 = 100$, and $\kappa_3 = 0.1^2$.

4.2. Sensitivity Results

Inference on the VAR model parameters is based on the output of 10,000 iterations (after the 1,000 burn-in iterations) of a 2-step Gibbs algorithm provided in Appendix B with posterior estimates of the model parameters provided in Appendix C. The IPA-based sensitivity analysis is again implemented via the vector-based AD approach also used in the simulation study (Kwok, Zhu, and Jacobi 2020). It investigates sensitivities of the estimated model parameters and GDP responses to government spending shocks in terms of FEVDs and IRFs with respect to the key prior parameters in the Minnesota context, the prior locations (β_0) and variances (\mathbf{B}_0) for the intercept and lag coefficients. In addition, we provide sensitivities for the companion matrix as a measure of the stability of a VAR system. In the final section we show how the introduced IPA estimates can be used to assess the impacts of prior parameter changes in IRFs across the complete forecast horizon.

4.2.1. Model Parameters and Model Stability

To begin with, we provide results on sensitivities of the posterior mean model parameter estimates across the three equations. Figure 7 presents heat maps of the mean derivatives for the model coefficients β for each equation (Government spending, GDP and Federal tax) with respect to prior means (top panel) and prior variances (bottom panel) of VAR parameters within the GDP equation. In the heatmap, each row corresponds to the derivatives of the posterior coefficient mean concerning all 13 prior mean/variance parameters in the GDP equation. For a more comprehensive illustration, please refer to Appendix E. Common across all panels, we see pronounced sensitivity patterns with respect to some prior settings for

¹¹These include real GDP (GDPC1), nominal government purchases, implicit price deflator (GDPDEF) and nominal Federal current receipts, NIPA accrual basis (FGRECPT).

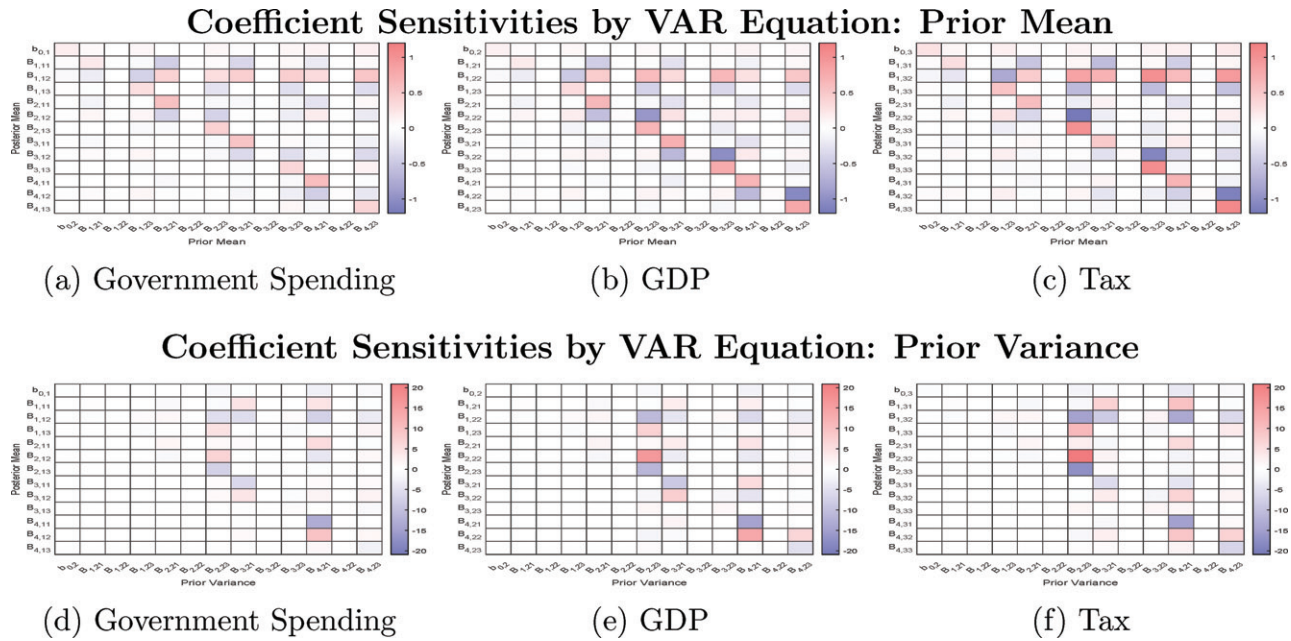


Figure 7. Mean derivatives of model coefficients from all three equations with respect to the prior means $\frac{\partial \hat{\beta}}{\partial \beta_0}$ (top panels) and prior variances $\frac{\partial \hat{\beta}}{\partial \text{diag}(\mathbf{B}_0)}$ (bottom panels) of prior parameters in the GDP equation.

Sensitivities of Companion Matrix

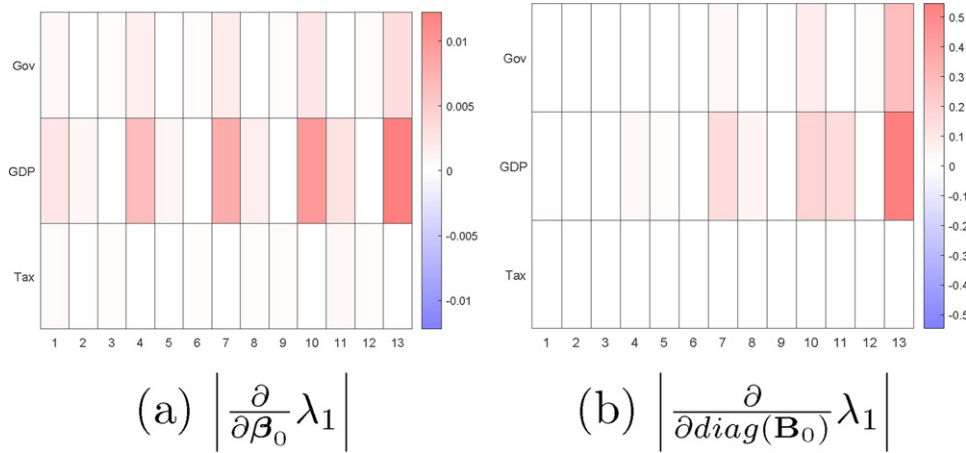


Figure 8. Heatmaps of derivative estimates (in absolute values) of largest eigenvalue λ_1 with respect to the prior mean and variance parameters. The three rows are derivatives of λ_1 with respects to prior parameters associated with the Government spending, GDP, and Tax equations, respectively.

parameters in the GDP equation where the heatmaps contain the darkest red (largest positive) or blue (most negative) values.

Moving forward, we next examine the sensitivities related to the largest absolute eigenvalue of the companion matrix, denoted as λ_1 . The companion matrix holds a pivotal role in ascertaining the stability characteristics of the VAR system under consideration. The stability of a VAR model is contingent upon all eigenvalues exhibiting an absolute value that is less than 1. This property ensures that the system's behavior remains non-explosive and permits the derivation of a unique, stationary representation. Furthermore, λ_1 , governs the speed at which the impacts of exogenous shocks within the system attenuate over time. Our exploration into the sensitivity of λ_1 concerning prior specifications yields valuable insights into the influence

of these priors on the long-term stability of the system. This becomes especially pertinent when λ_1 converges toward a value of 1. In the aforementioned analysis, it has been observed that the model parameters are most sensitive to the prior means and variances related to the GDP equation. This pattern is further illustrated in Figure 8, where the absolute values of the IPA estimates of the largest eigenvalue of the companion matrix, λ_1 , are displayed. It is noteworthy that the absolute values of the posterior means for λ_1 are slightly below 1, indicating proximity to this threshold value. Given the existing prior parameter configurations, the estimated VAR model aligns closely with a unit root process. However, should the prior variances for the lag coefficients associated with Federal Tax in the GDP equation experience an increase, it is likely that λ_1 would surpass the value of 1.

Sensitivities of IRF and FEVD for GDP at 20 quarters

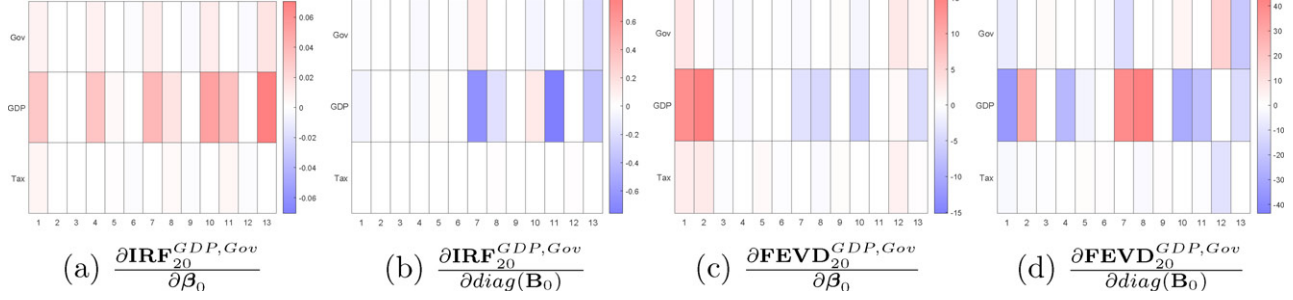


Figure 9. Heatmaps of derivative estimates of the IRF and FEVD of log GDP at 20 quarters after the spending shock with respect to the prior mean parameters and variance parameters. The three rows are derivatives with respects to prior parameters associated with the Government spending, GDP and Tax equations, respectively.

4.2.2. FEVD and IRF of GDP to Spending Shock

In the of the section, we focus on GDP responses from an orthogonal spending shock as implied by the Cholsky decomposition. Figure 9 presents the sensitivity estimates concerning the 20-quarter ahead IRFs and FEVD of the log GDP in response to a government spending shock, denoted as $\text{IRF}_{20}^{GDP,Gov}$ and $\text{FEVD}_{20}^{GDP,Gov}$, respectively. Each heatmap in these graphs encompasses 39 derivative values relative to the prior mean or variance parameters. We note that the majority of sensitivities are strongly linked to the prior means and variances in the GDP equation, which is the consistent with the results presented in the previous section. Moreover, it's worth noting that the intercept term in the GDP equation holds significant importance for the $\text{FEVD}_{20}^{GDP,Gov}$.

4.2.3. Prior Parameter Dependence across the IRF Forecast Horizon

Consider a forecast horizon of $T_2 = 24$. We use the notation $\hat{\text{IRF}}_t^{GDP,Gov}(\eta_0)$ to represent the estimated IRFs for logarithmic GDP under a government spending shock for time periods $t = 1, \dots, T_2$, given a vector of hyperparameters η_0 . Correspondingly, $\nabla \hat{\text{IRF}}_t^{GDP,Gov}(\eta_0)$ denotes the corresponding estimate of derivatives for $\eta_0 = (\beta_0, \text{diag}(\mathbf{B}_0))$. It is important to recognize that it is feasible to calculate derivatives of

$$\hat{\text{IRF}}_t^{ij}(\eta_0) \forall i \in [1, \dots, n], j \in [1, \dots, n], t \in [1, \dots, T_2]$$

simultaneously, that is, obtain the matrix $\nabla \hat{\text{IRF}}(\eta_0)$ of dimensions $n^2 T_2 \times \dim(\eta_0)$.

By making small perturbations to the prior parameters and observing the resulting changes in the IRFs, analysts can acquire valuable insights into the robustness of their chosen prior specifications. In this section, our objective is to showcase the use of IPA derivatives in efficiently conducting such analyzes. This obviates the need for repetitive runs of MCMC inference at various perturbed values of the prior hyperparameters. To do so, we consider a first-order Taylor approximation at each point in time

$$\hat{\text{IRF}}_t^{GDP,Gov}(\eta_0 + \delta) \approx \hat{\text{IRF}}_t^{GDP,Gov}(\eta_0) + \nabla \hat{\text{IRF}}_t^{GDP,Gov}(\eta_0)' \delta \quad (7)$$

for a basic perturbation analysis on the (sub)set of prior parameters, with the perturbation direction δ . While a single run of MCMC provides estimations for $\hat{\text{IRF}}_t^{GDP,Gov}(\eta_0)$

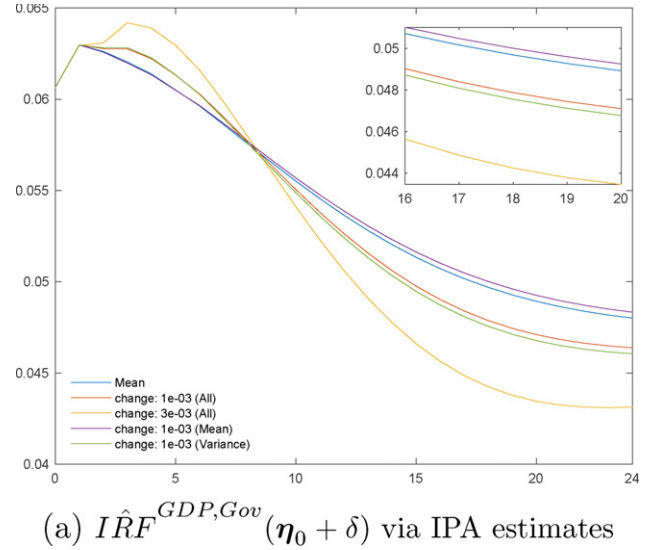


Figure 10. IPA-based GDP response changes ($\hat{\text{IRF}}_t^{GDP,Gov}(\eta_0 + \delta)$) under various perturbations ($\delta = 1e-03, \delta = 3e-03$) of prior parameter settings (prior mean and variance parameters, prior mean parameters, prior variance parameters) under RW prior.

and $+\nabla \hat{\text{IRF}}_t^{GDP,Gov}(\eta_0)$, we can extend this to derive various $\hat{\text{IRF}}_t^{GDP,Gov}(\eta_0 + \delta)$ estimates by evaluating the outlined approximation at different values of δ .

In order to assess the overall prior sensitivity of the IRFs of log GDP with respect to a government spending shock, we consider a very small change in η_0 on mean estimates of the IRF. Figure 10 presents the IRF under the initial specification of the prior location and variance parameters, alongside the implied IRF for 24 periods following the shock when either the prior means, prior variances or both are changed by $1e-3$ (and also $3e-3$ for the latter). Two key patterns emerge. First, the highest sensitivities are observed in the responses further out in the time horizon, particularly between 15 and 20 quarters after the spending shock. Second, responses are more sensitive to the changes in the prior variance parameters, which is expected given their shrinkage effect and their lower magnitudes.

The accuracy of such first-order approximations is influenced by the curvature of the function being approximated. The curvature provides insights into how the function's rate of change varies across different directions. The curvature is captured by the Hessian matrix, which contains second-order

IRF Changes under IPA-Approximation and Re-Estimation

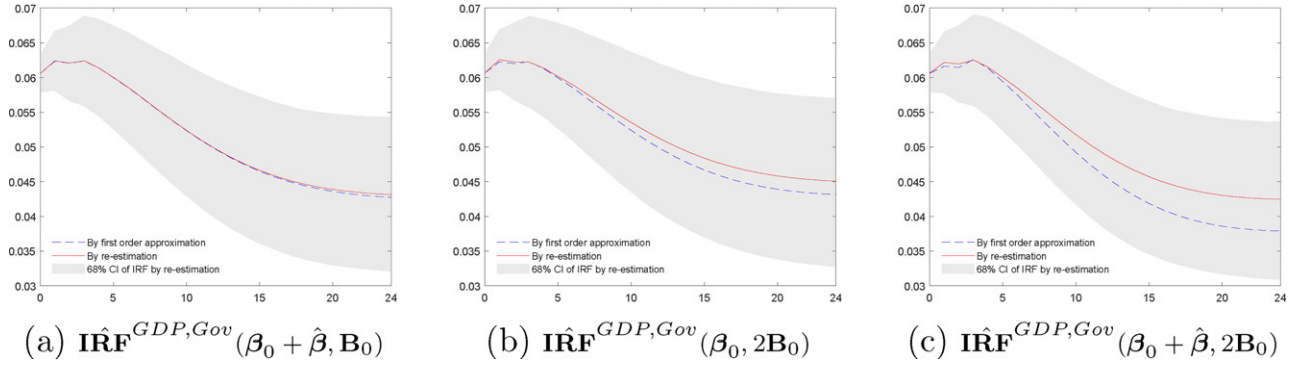


Figure 11. Approximated change of GDP response to spending shock, $\hat{\text{IRF}}^{GDP, Gov}$, under relative changes in prior mean and variances based on AD derivatives/first order Taylor series expansion compared to 68% IRF interval under initial prior setting.

IPA-Approximated IRF Under Discrete Prior Parameter Changes

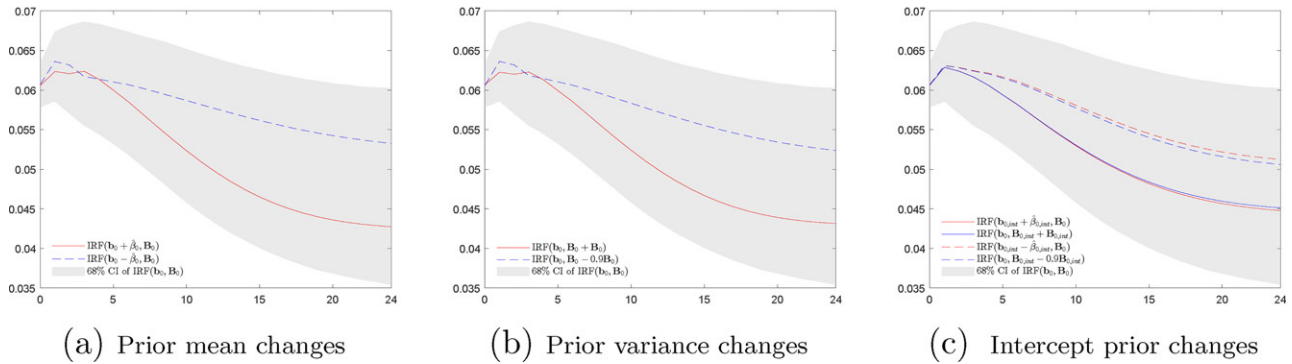


Figure 12. Approximated change of GDP response to spending shock, $\hat{\text{IRF}}^{GDP, Gov}$, under relative changes in prior mean and variances based on AD derivatives/first order Taylor series expansion compared to 68% IRF interval under initial prior setting. For changes in the intercept prior both mean and variances are changed.

partial derivatives of the function with respect to its variables. However, computing the Hessian matrix can be computationally demanding and time-consuming, particularly for complex functions or high-dimensional problems.

Here, we present the outcomes of our first-order approximated Impulse Response Function (IRF) through IPA derivatives in comparison to the IRF obtained by conducting a further run of the Gibbs sampler using perturbed prior hyperparameters. The graphs illustrated in Figure 11 demonstrate that across all considered modifications, the IRF estimates derived from the derivative-based approach are either nearly indistinguishable or extremely close to the IRF acquired through re-estimation with the adjusted prior values. These approximations consistently fall well within the 68% credibility intervals. As anticipated, the approximation exhibits the highest accuracy for changes in means due to the linear nature of IRF with respect to the means. Conversely, the presence of non-linearity within the variance parameters leads to a slightly larger approximation error. This highlights that the influence of second-order terms cannot be disregarded, particularly when dealing with non-linearities stemming from variance parameters.

Finally, within Figure 12, we further analyse how the prior specification of intercepts influences IRFs. It is important to note that while the intercept itself does not have a direct impact on IRF calculations, its prior effects are inherently embedded

within IRFs through its posterior interdependence with other model parameters. The panels (a) and (b) of the figure examine scenarios involving increases and decreases in prior mean or variance by their posterior estimates, respectively. The concluding panel of the figure delves into adjustments made to both the prior mean and variance parameters associated with the intercept terms. This comprehensive exploration enables us to understand how modifications to the prior assumptions for intercepts reverberate through the IRF estimations, considering the intricate interplay between these parameters and other components of the model. The results suggest that IRF estimates are quite robust to changes in the prior parameters at least in this particular context. That is, with different changes, the IRF estimates are still well within the IRF interval estimated under the original settings.

5. Discussion

The dependence of posterior inference on a (often large) set of prior parameters, typically specified by the researcher, is an inherent feature of Bayesian empirical analysis. Local sensitivity analysis can provide important insights into the patterns of prior parameter dependence for key posterior statistics. However, the computation of such prior parameter sensitivities is complicated by statistical and computational challenges and available

methods are limited in scope and computational feasibility. This article introduces a new approach based on Infinitesimal Perturbation Analysis to expand the scope of sensitivity analysis.

With IPA widely used in classical simulation, we show that it can be extended to compute asymptotically unbiased and consistent prior parameter sensitivities of posterior statistics obtained from Markov chain Monte Carlo inference via Gibbs sampling under set of regularity conditions. By decomposing the Gibbs sampler into a sequence of mappings, we are able to compute the derivatives estimate alongside the MCMC algorithm via the automatic differentiation, that avoid repeated rerunning the MCMC. The new method also provides a more general approach to unbiased derivative estimation than the LR method that is applicable to a wider range of statistics computed from the model parameters, and does not suffer from the computational and statistical drawbacks of the FD methods. However, it is important to acknowledge that the IPA approach has a potential limitation due to the required smoothness in the underlying algorithm. Some widely-used sampling methods beyond Gibbs, such as unadjusted Hamiltonian Monte Carlo and Langevin sampling, adhere to these smoothness prerequisites. Accept-reject type algorithms such as Metropolis Hastings fail this condition and an extension of IPA estimates requires further work. We illustrate how IPA might applied in this type of setting in the context of Metropolis-within-Gibbs and Slice-within-Gibbs algorithms. A more comprehensive examination of sensitivity analysis for MCMC outputs under Metropolis-Hasting type algorithms is a matter for subsequent research.

We illustrate the performance and relevance of the proposed approach in the context of Bayesian VAR analysis where we assess the dependence of VAR inference to prior location and variance specifications under a common Minnesota-type prior. The simulation study confirms the unbiasedness and precision in the benchmarking against the LR and FD methods. It also introduces IPA-based derivatives for posterior statistics beyond posterior parameter estimates, to include the Impulse Response Functions and Forecast Error Variance Decompositions of orthogonal exogenous shocks. Our application in terms of a fiscal policy-VAR for U.S. data (Government Spending, GDP, Tax) provides the first formal local sensitivity analysis of IRFs and finds the highest sensitivities with respect to prior location and variance parameters related to the GDP equation, a finding that consistent across all posterior measures.

Supplementary Materials

This supplementary material comprises comprehensive proofs of the theoretical results, accompanied by supplementary simulation data and empirical findings.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Mark Joshi  <http://orcid.org/0000-0002-8456-3438>

References

- Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679. [135]
- Amir-Ahmadi, P., Matthes, C., and Wang, M.-C. (2018), "Choosing Prior Hyperparameters: With Applications to Time-Varying Parameter Models," *Journal of Business & Economic Statistics*, 38, 124–136. [134,143]
- Baumeister, C., and Hamilton, J. D. (2019), "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks," *American Economic Review*, 109, 1873–1910. [134]
- Baumeister, C., Korobilis, D., and Lee, T. K. (2020), "Energy Markets and Global Economic Conditions," *The Review of Economics and Statistics*, 104, 1–45. [134]
- Berger, J. O., Insua, D. R., and Ruggeri, F. (2000), "Bayesian Robustness," in *Robust Bayesian Analysis*, eds. D. R. Insua and F. Ruggeri, pp. 1–32, New York: Springer. [134,135,136]
- Blanchard, O., and Perotti, R. (2002), "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *the Quarterly Journal of Economics*, 117, 1329–1368. [143]
- Bou-Rabee, N., Eberle, A., and Zimmer, R. (2020), "Coupling and Convergence for Hamiltonian Monte Carlo," *The Annals of Applied Probability*, 30, 1209–1250. [138]
- Briggs, A. D., Mytton, O. T., Kehlbacher, A., Tiffin, R., Elhussein, A., Rayner, M., Jebb, S. A., Blakely, T., and Scarborough, P. (2017), "Health Impact Assessment of the UK Soft Drinks Industry Levy: A Comparative Risk Assessment Modelling Study," *The Lancet Public Health*, 2, e15–e22. [135]
- Camacho, M., Gadea, M. D., and Gómez-Loscos, A. (2020), "A New Approach to Dating the Reference Cycle," *Journal of Business & Economic Statistics*, 40, 66–81. [135]
- Carriero, A., Clark, T. E., and Marcellino, M. (2019), "Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *Journal of Econometrics*, 212, 137–154. [139]
- Carter, C. K., and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553. [138]
- Chib, S., and Ergashev, B. (2009), "Analysis of Multifactor Affine Yield Curve Models," *Journal of the American Statistical Association*, 104, 1324–1337. [136]
- Cui, Z., Fu, M. C., Hu, J.-Q., Liu, Y., Peng, Y., and Zhu, L. (2020), "On the Variance of Single-Run Unbiased Stochastic Derivative Estimators," *INFORMS Journal on Computing*, 32, 390–407. [140]
- De Jong, R. M. (1995), "Laws of Large Numbers for Dependent Heterogeneous Processes," *Econometric Theory*, 11, 347–358. [138]
- Del Negro, M., and Schorfheide, F. (2008), "Forming Priors for DSGE Models (and How it Affects the Assessment of Nominal Rigidities)," *Journal of Monetary Economics*, 55, 1191–1208. [134]
- Diebold, F. X., and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337–364. [138]
- Doan, T., Litterman, R., and Sims, C. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100. [135]
- Durbin, J., and Koopman, S. J. (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–616. [135]
- Durmus, A., and Moulines, E. (2017), "Nonasymptotic Convergence Analysis for the Unadjusted Langevin Algorithm," *The Annals of Applied Probability*, 27, 1551–1587. [138]
- Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202. [135]
- Giannone, D., Lenza, M., and Primiceri, G. E. (2015), "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97, 436–451. [134]
- Giles, M., and Glasserman, P. (2006), "Smoking Adjoints: Fast Monte Carlo Greeks," *Risk*, 19, 88–92. [136]
- Giordano, R., Broderick, T., and Jordan, M. I. (2017), "Covariances, Robustness, and Variational Bayes," arXiv preprint arXiv:1709.02536. [136,139]
- Glasserman, P. (2013), *Monte Carlo Methods in Financial Engineering* (Vol. 53), New York: Springer. [135,139,140,141]

- Glasserman, P., and Liu, Z. (2010), "Estimating Greeks in Simulating lévy-driven Models," *Journal of Computational Finance*, 14, 3. [136]
- Greenberg, E. (2012), *Introduction to Bayesian Econometrics*, Cambridge: Cambridge University Press. [134]
- Griewank, A., and Walther, A. (2008), *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation* (Vol. 105), Philadelphia: SIAM. [135,138]
- Gustafson, P. (2000), "Local Robustness in Bayesian Analysis," in *Robust Bayesian Analysis*, Springer, pp. 71–88. [134,135]
- Hauzenberger, N., Huber, F., Marcellino, M., and Petz, N. (2021), "Gaussian Process Vector Autoregressions and Macroeconomic Uncertainty," arXiv preprint arXiv:2112.01995. [143]
- Homescu, C. (2011), "Adjoint and Automatic (Algorithmic) Differentiation in Computational Finance," Available at SSRN: <https://doi.org/10.2139/ssrn.1828503>. [138]
- Huber, F., and Feldkircher, M. (2019), "Adaptive Shrinkage in Bayesian Vector Autoregressive Models," *Journal of Business & Economic Statistics*, 37, 27–39. [134]
- Jacobi, L., Nghiem, N., Ramirez Hassan, A., and Blakely, T. (2021), "Food Price Elasticities for Policy Interventions: Estimates from a Virtual Experiment in a Multistage Demand Analysis with (Expert) Prior Information," *Economic Record* (forthcoming). [134]
- Jacobi, L., Wagner, H., and Frühwirth-Schnatter, S. (2016), "Bayesian Treatment Effects Models with Variable Selection for Panel Outcomes with an Application to Earnings Effects of Maternity Leave," *Journal of Econometrics*, 193, 234–250. [135]
- Jarociński, M., and Marcet, A. (2019), "Priors about Observables in Vector Autoregressions," *Journal of Econometrics*, 209, 238–255. [134,143,144]
- Karlsson, S. (2013), "Forecasting with Bayesian Vector Autoregression," in *Handbook of Economic Forecasting* (Vol. 2), Elsevier, pp. 791–897. [139,144]
- Koop, G., and Korobilis, D. (2010), *Bayesian Multivariate Time Series Methods for Empirical Macroeconomics*, Delft: Now Publishers Inc. [135,139,144]
- Kwok, C. F., Zhu, D., and Jacobi, L. (2020), *ADtools: Automatic Differentiation Toolbox. R package version 0.5.4*, <https://cran.r-project.org/package=ADtools>. [139,144]
- (2022), "An Analysis of Vectorised Automatic Differentiation for Statistical Applications," Available at SSRN. [139]
- L'Ecuyer, P. (1990), "A Unified View of the ipa, sf, and Ir Gradient Estimation Techniques," *Management Science*, 36, 1364–1383. [137]
- Litterman, R. B. (1986), "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience," *Journal of Business & Economic Statistics*, 4, 25–38. [135]
- Morley, J., and Wong, B. (2017), "Estimating and Accounting for the Output Gap with Large Bayesian Vector Autoregressions," *CAMA working paper*, 46. [139]
- Müller, U. K. (2012), "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *Journal of Monetary Economics*, 59, 581–597. [135,136,139]
- Peng, Y., Fu, M. C., Hu, J.-Q., and Heidergott, B. (2018), "A New Unbiased Stochastic Derivative Estimator for Discontinuous Sample Performances with Structural Parameters," *Operations Research*, 66, 487–499. [139]
- Pérez, C., Martin, J., and Rufo, M. (2006), "MCMC-based Local Parametric Sensitivity Estimations," *Computational Statistics & Data Analysis*, 51, 823–835. [135,136,139]
- Primiceri, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852. [138]
- Raghunathan, M. S. (1979), "A Proof of Oseledec's Multiplicative Ergodic Theorem," *Israel Journal of Mathematics*, 32, 356–362. [138]
- Ramey, V. A. (2019), "Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives*, 33, 89–114. [135,143,144]
- Roos, M., Martins, T. G., Held, L., Rue, H., et al. (2015), "Sensitivity Analysis for Bayesian Hierarchical Models," *Bayesian Analysis*, 10, 321–349. [136]
- Townsend, J., Koep, N., and Weichwald, S. (2016), "Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation," *The Journal of Machine Learning Research*, 17, 4755–4759. [139]