

DISCUSSION PAPER SERIES

DP17646

GAUSSIAN PROCESS VECTOR AUTOREGRESSIONS AND MACROECONOMIC UNCERTAINTY

Niko Hauzenberger, Florian Huber, Massimiliano
Marcellino and Nico Petz

**MONETARY ECONOMICS AND
FLUCTUATIONS**

CEPR

GAUSSIAN PROCESS VECTOR AUTOREGRESSIONS AND MACROECONOMIC UNCERTAINTY

Niko Hauzenberger, Florian Huber, Massimiliano Marcellino and Nico Petz

Discussion Paper DP17646
Published 04 November 2022
Submitted 03 November 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Niko Hauzenberger, Florian Huber, Massimiliano Marcellino and Nico Petz

GAUSSIAN PROCESS VECTOR AUTOREGRESSIONS AND MACROECONOMIC UNCERTAINTY

Abstract

We develop a non-parametric multivariate time series model that remains agnostic on the precise relationship between a (possibly) large set of macroeconomic time series and their lagged values. The main building block of our model is a Gaussian process prior on the functional relationship that determines the conditional mean of the model, hence the name of Gaussian process vector autoregression (GP-VAR). A flexible stochastic volatility specification is used to provide additional flexibility and control for heteroskedasticity. Markov chain Monte Carlo (MCMC) estimation is carried out through an efficient and scalable algorithm which can handle large models. The GP-VAR is illustrated by means of simulated data and in a forecasting exercise with US data. Moreover, we use the GP-VAR to analyze the effects of macroeconomic uncertainty, with a particular emphasis on time variation and asymmetries in the transmission mechanisms.

JEL Classification: C11, C14, C32

Keywords: N/A

Niko Hauzenberger - niko.hauzenberger@plus.ac.at
University of Salzburg

Florian Huber - florian.huber@plus.ac.at
University of Salzburg

Massimiliano Marcellino - massimiliano.marcellino@unibocconi.it
Bocconi University and CEPR

Nico Petz - nico.petz@plus.ac.at
University of Salzburg

Gaussian Process Vector Autoregressions and Macroeconomic Uncertainty*

Niko HAUZENBERGER^a, Florian HUBER^a, Massimiliano MARCELLINO^b, and Nico PETZ^a

^a*University of Salzburg*

^b*Bocconi University, IGIER, CEPR, Baffi-Carefin and BIDS*

November 3, 2022

Abstract. We develop a non-parametric multivariate time series model that remains agnostic on the precise relationship between a (possibly) large set of macroeconomic time series and their lagged values. The main building block of our model is a Gaussian process prior on the functional relationship that determines the conditional mean of the model, hence the name of Gaussian process vector autoregression (GP-VAR). A flexible stochastic volatility specification is used to provide additional flexibility and control for heteroskedasticity. Markov chain Monte Carlo (MCMC) estimation is carried out through an efficient and scalable algorithm which can handle large models. The GP-VAR is illustrated by means of simulated data and in a forecasting exercise with US data. Moreover, we use the GP-VAR to analyze the effects of macroeconomic uncertainty, with a particular emphasis on time variation and asymmetries in the transmission mechanisms.

JEL: C11, C14, C32, E32.

KEYWORDS: Bayesian non-parametrics, non-linear vector autoregressions, asymmetric uncertainty shocks.

* *Corresponding author:* Massimiliano Marcellino. Bocconi University. *Address:* Via Roentgen 1, I-20136 Milano. *Email:* massimiliano.marcellino@unibocconi.it. We would like to thank for useful comments on a previous draft Roberto Casarin, Karim Chalak, Joshua Chan, Todd Clark, Sharada Nia Davidson, Jean-Marie Dufour, Raffaella Giacomini, William McCausland, Serena Ng, Luca Onorante, Michael Pfarrhofer, Barbara Rossi, Samad Sarferaz, Frank Schorfheide, Anna Stelzer, Dalibor Stevanovic, Tommaso Tornese, an associate editor as well as two referees and the participants of the 8th Annual Conference of the International Association for Applied Econometrics (IAAE), the 12th European Seminar on Bayesian Econometrics (ESOB), the EABCN and Bundesbank conference on challenges in empirical macroeconomics since 2020, the Decision Making under Uncertainty (DeMUr) 2022 workshop and the University of Montreal econometrics seminar. Hauzenberger gratefully acknowledges financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (OeNB, grant no. 18304 and 18718), and Huber also acknowledges financial support from the Austrian Science Fund (FWF, grant no. ZK 35) and the Jubiläumsfonds of the OeNB (grant no. 18304).

1 Introduction

Economic relations can change over time for a variety of reasons, such as technological progress, institutional changes, major policy interventions, but also wars, terrorist attacks, stock market crashes and pandemics. Standard econometric models, such as linear single and multivariate regressions assume instead stability of the parameters characterizing the conditional first and second moments of the dependent variables. When stability is formally tested, it is often rejected (see, e.g., [Stock and Watson, 1996](#)). This has led to the development of a variety of methods to handle structural change in econometric models.

Parameter evolution is assumed to be either observable (i.e., driven by the behavior of observable economic variables) or unobservable, and either discrete and abrupt or continuous and smooth. Examples include threshold and smooth transition models (see, e.g., [Tong, 1990](#); [Teräsvirta, 1994](#)), Markov switching models (see, e.g., [Hamilton, 1989](#)), and double stochastic models (see, e.g., [Nyblom, 1989](#)). In all these models, a specific and fully parametrized type of parameter evolution is assumed, and then linear or non-linear filters are used for estimation in a classical context or Markov chain Monte Carlo (MCMC) methods in a Bayesian framework. Examples of economic applications of all these methods include [Koop and Korobilis \(2013\)](#), [Aastveit, et al. \(2017\)](#), [Aastveit, Natvik, and Sola \(2017\)](#), [Caggiano, Castelnovo, and Pellegrino \(2017\)](#), [Alessandri and Mumtaz \(2019\)](#), and [Caggiano, Castelnovo, and Pellegrino \(2021\)](#).¹

Assuming a specific type of parameter evolution increases estimation efficiency but can lead to mis-specification. A more flexible alternative allows for a smooth evolution of parameters without specifying the form of parameter time variation. In a classical context, the evolution can be either deterministic (see, e.g., [Robinson, 1991](#); [Chen and Hong, 2012](#)), or stochastic (see, e.g., [Giraitis, Kapetanios, and Yates, 2014, 2018](#); [Kapetanios, Marcellino, and Venditti, 2019](#), for the specific case of (possibly large) vector autoregressive models (VARs)). Kernel estimators are the main tool used in this literature. Alternative approaches, which can also capture non-linear relationships between the target and explanatory variables, are, e.g., based on using functional-coefficient regressions ([Cai, Fan, and Yao, 2000](#); [Kowal, Matteson, and Ruppert, 2017](#)), regression trees ([Chipman, George, and McCulloch, 2010](#); [Huber, et al., 2020](#); [Coulombe, 2020](#)), neural networks

¹A special mention is due to [Primiceri \(2005\)](#) who popularized the use of time-varying parameters and stochastic volatility in macroeconometrics.

(Hornik, Stinchcombe, and White, 1989; Gu, Kelly, and Xiu, 2021; Coulombe, 2022) or infinite mixtures (Hirano, 2002; Bassetti, Casarin, and Leisen, 2014; Kalli and Griffin, 2018; Billio, Casarin, and Rossini, 2019; Jin, Maheu, and Yang, 2022). Most of these approaches, however, are fairly different from the VAR models that are the workhorse of modern time series econometrics, making interpretation of the estimation results and computation of quantities such as impulse response functions difficult. In addition, they typically focus on extending the specification of the conditional mean, while assuming a constant conditional variance, which can be restrictive for macroeconomic and financial data. Finally, some of the methods do not scale well into high dimensions and are thus not particularly suited for large datasets nowadays used in macroeconomics.

In this paper, we propose a new model that belongs to the non-parametric class and is capable of capturing, in a very flexible way, not only parameter evolution but also general non-linear relationships. Our model can be applied in a large data context while keeping the flexibility and ease of use of VARs, and allowing for time-varying conditional variances. Specifically, we combine the statistical literature on Gaussian process (GP) regressions (see, e.g., Crawford, et al., 2019), with that on VARs to obtain a GP-VAR model. Borrowing ideas from the literature on Bayesian Minnesota-type VARs, the model assumes, for each endogenous variable, a different non-linear relationship with its own lags and with the lags of all the other variables (and possibly of additional exogenous regressors). Gaussian processes are used to model non-linearities in a flexible but efficient way. They can be viewed as a non-parametric alternative to the adaptive Minnesota-type shrinkage proposed in Chan (2021). They are also similar to neural networks, in the sense that they are universal approximators based on infinite mixtures of Gaussian distributions.²

We develop an efficient (Bayesian) estimation procedure, based on the structural form of the GP-VAR, which has the additional benefit that its complexity is linear in the number of endogenous variables and does not depend on the number of lags. Hence, estimation can be parallelized and, in addition, the conjugate structure of the model we use allows for pre-computing various matrix multiplications and kernel operations, which further speeds up computation. As a result, estimation is feasible also for very large models.

²In fact, specific choices of the kernels underlying Gaussian processes can produce a variety of neural network models, see Novak, et al. (2018) for details.

As in all Bayesian procedures, an assumption on the distribution of the errors is required. As is common in the Bayesian VAR literature, we assume that the errors are Gaussian. Yet, we also permit the variance of the errors to change over time, adopting a stochastic volatility (SV) specification. Already in linear BVARs, the use of SV permits to have time variation in the conditional distribution of the variables, so much so that BVAR-SV are empirically a good alternative to quantile regressions (see [Carriero, Clark, and Marcellino, 2022](#)). Moreover, multi-step predictive densities, used to produce forecasts and generalized impulse responses, are non-Gaussian. The use of SV in the GP-VAR adds flexibility, and prevents overfitting in the sense of avoiding that large realizations of the shocks are interpreted as changes in the conditional mean.

We illustrate the GP-VAR with synthetic data generated from a highly non-linear multivariate data generating process (DGP) that have both Gaussian and non-Gaussian shocks. This DGP assumes that some equations feature parameters that exhibit structural breaks while others depend non-linearly on the lags of the endogenous variables. To assess whether the GP-VAR is also capable of recovering linear relations, one equation is a standard linear regression model. In all these cases, our approach works reasonably well in terms of detecting the nature of non-linearities.

Our GP-based model has a vast range of applicability, for both reduced form and structural analysis. This mirrors the possible applications of standard VARs but allows for much more general dynamic relationships across the variables. Besides the evaluation with synthetic data, we consider a forecasting application and a more structural economic analysis. In the forecasting application, we compare the performance of the GP-VAR with other linear and non-linear competitors that allow for parameter change, and with the BVAR with SV. Predicting US output, inflation and interest rates, we show that the GP-VAR improves upon all competing models, with gains that are particularly pronounced at the four-quarter-ahead horizon.

As an example of a structural economic analysis, and to gather new insights on a topic that has recently attracted considerable attention (see, e.g., [Bloom, 2014](#)), we use the GP-VAR to investigate the effects of exogenous uncertainty shocks on US macroeconomic and financial time series. Comparing the responses of the GP-VAR with the ones of a standard linear BVAR reveals that our model produces sensible responses for real activity and stock markets. Differences between the responses relate to the shape and magnitudes,

with the GP-VAR producing stronger reactions of uncertainty, real GDP growth, and stock markets returns while yielding similar reactions of employment growth. Considering models of differing sizes shows that the impulse responses do not differ markedly across model sizes.

Our proposed framework naturally allows for analyzing potential asymmetries in transmission channels. The responses to a positive uncertainty shock (higher unexpected uncertainty) are typically much stronger than the ones to a negative shock. Interestingly, the shape of the IRFs also differ, with positive shocks leading to responses that peak later. In addition, our findings suggest that the relationship between real activity and uncertainty becomes proportionally slightly smaller for large shocks, while financial markets react relatively more strongly to larger increases in uncertainty. Finally, our framework also allows us to investigate whether transmission mechanisms have changed over time. Doing so reveals that the effects of uncertainty have been smaller in the great inflation period (1970Q1 to 1984Q4), more pronounced through the great moderation (1985Q1 to 2006Q4) before again turning more muted during the post great moderation phase (2007Q1 to 2019Q4).

The paper is structured as follows. Section 2 provides an introduction to Gaussian process regression. In Section 3 we develop the GP-VAR model. This section also provides necessary details on the prior setup and posterior computation. We then analyze model performance using synthetic data in Section 4. Sections 5 and 6 include our empirical work. In the former section we briefly discuss the dataset and provide some in- and out-of-sample model evidence. In the latter, we focus on the macroeconomic implications of an uncertainty shock. The final section briefly summarizes and concludes the paper. The Online Appendix contains additional empirical results and technical details on the specification, estimation and use of the GP-VAR model.

2 A brief introduction to Gaussian processes

In this section we briefly discuss Gaussian process (GP) regressions with a focus on time series data.³ GP regressions are a non-parametric technique to establish a flexible relationship between a scalar time series y_t and a set of K predictors \mathbf{x}_t in period t . The

³For a textbook treatment, see Williams and Rasmussen (2006).

key advantage of this approach is that it does not rely on parametric assumptions on the precise functional relationship between y_t and \mathbf{x}_t .

In general, a non-parametric regression is given by:

$$y_t = f(\mathbf{x}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2),$$

with f being some unknown regression function $f : \mathbb{R}^K \rightarrow \mathbb{R}$ and ε_t denoting an independent Gaussian shock with zero mean and constant variance σ^2 . We relax this assumption in Sub-section 3.1 to allow for heteroskedastic shocks. An assumption on the error distribution is needed in a Bayesian context and Gaussianity is the most common one, though different distributions can be easily accommodated by exploiting a scale-location mixture of Gaussians representation (see, e.g., Escobar and West, 1995).

In standard regression models, the function f is assumed to be linear with $f(\mathbf{x}_t) = \boldsymbol{\beta}'\mathbf{x}_t$ where $\boldsymbol{\beta}$ is a $K \times 1$ vector of linear coefficients. If mean relations are non-linear, this assumption might be too restrictive. To gain more flexibility one can embed the covariates in \mathbf{x}_t into a higher dimensional space such as the space of powers $\mathbf{x}_t \rightarrow \psi(\mathbf{x}_t) = (\mathbf{x}_t', (\mathbf{x}_t^2)', \dots, (\mathbf{x}_t^R)')'$, with $\mathbf{x}_t^2 = (\mathbf{x}_t \odot \mathbf{x}_t)$ and higher orders defined recursively. Conditional on choosing a sufficiently large integer R , this would provide substantial flexibility to approximate any smooth function f . However, adequately selecting R is key and the mapping, moreover, is ad-hoc in the sense that there exist infinitely many non-linear mappings ψ .

Standard Bayesian methods place a prior on the coefficients associated with the covariates (and possible non-linear transformations thereof) and thus control for uncertainty with respect to these basis functions but at the cost of remaining within a class of functions (such as linear, polynomial or trigonometric functions). By contrast, in GP regressions we treat the function f as an unknown quantity and let the data decide about the appropriate form (and degree) of non-linearities.

2.1 Estimating unknown functions: the function space view

The key inferential goal in GP regression is to infer the function f from the data under relatively mild assumptions. This is achieved by specifying a prior on $f(\mathbf{x}_t)$. A typical

assumption is to assume that $f(\mathbf{x}_t)$ follows a Gaussian process prior:

$$f(\mathbf{x}_t) \sim \mathcal{GP}(\mu(\mathbf{x}_t), k_{\boldsymbol{\vartheta}}(\mathbf{x}_t, \mathbf{x}_t)),$$

with $\mu(\mathbf{x}_t) = \mathbb{E}[f(\mathbf{x}_t)]$ being the mean function and

$$k_{\boldsymbol{\vartheta}}(\mathbf{x}_t, \mathbf{x}_\tau) = \mathbb{E}[(f(\mathbf{x}_t) - \mu(\mathbf{x}_t))(f(\mathbf{x}_\tau) - \mu(\mathbf{x}_\tau))]$$

denoting a kernel (or covariance) function that determines the relationship between $f(\mathbf{x}_t)$ and $f(\mathbf{x}_\tau)$ for periods t and τ . The kernel is typically parameterized by a low dimensional vector of hyperparameters $\boldsymbol{\vartheta}$ and controls the behavior of the function f . This kernel needs to be positive semidefinite and symmetric.

In what follows, we will set the function $\mu(\mathbf{x}_t) = 0$ for all t . This is without loss of generality, since any explicit basis function for $\mu(\mathbf{x}_t)$ can be used to model the mean process. If the focus is on modeling stationary data, $\mu(\mathbf{x}_t) = 0$ implies that a priori the process is centered around a white noise process. In case one would like to model persistent or non-stationary data it would be straightforward to implement a prior that forces the system towards a set of random walk processes. This can be achieved by setting $\mu(y_{t-1}) = \rho y_{t-1}$, where ρ denotes a persistence parameter with prior mean $\mathbb{E}[\rho] = 1$. Alternatively, one could specify the prior on f to imply persistence in y_t . This possibility is discussed in much more detail in Section A.1 of the Online Appendix.

A common choice in GP regressions is the Gaussian (or squared exponential) kernel function:

$$k_{\boldsymbol{\vartheta}}(\mathbf{x}_t, \mathbf{x}_\tau) = \xi \times \exp\left(-\frac{\kappa}{2}\|\mathbf{x}_t - \mathbf{x}_\tau\|^2\right),$$

with ξ denoting a scaling parameter and κ the (inverse) length scale and thus $\boldsymbol{\vartheta} = (\xi, \kappa)'$. Larger values of κ lead to a GP which displays more high frequency variation whereas lower values imply a slowly varying mean function. The parameter ξ controls the prior variance of the function f . To see this, note that if $\mathbf{x}_t = \mathbf{x}_\tau$, we obtain $\text{Var}[f(\mathbf{x}_t)] = \xi$.

This specification is quite flexible and fulfills several convenient conditions. For instance, Williams and Rasmussen (2006) show that the use of the Gaussian kernel implies that $f(\mathbf{x}_t)$ is mean square continuous and differentiable. Moreover, this kernel function

represents a positive semidefinite and symmetric covariance function. Furthermore, Mercer’s theorem (Mercer, 1909), under this kernel, states that the GP regression can be written in terms of an infinite number of basis functions. These basis functions are Gaussians with different means and variances. This suggests a connection to the literature on Bayesian non-parametrics (Escobar and West, 1995; Neal, 2000; Kalli and Griffin, 2018; Frühwirth-Schnatter and Malsiner-Walli, 2019) that relies on infinite mixtures of Gaussians to estimate unknown densities. The link between GPs and infinite mixtures of Gaussians shows that the Gaussian assumption on ε_t is not too restrictive as the model allows to recover non-Gaussian features in the data.

The GP prior represents an infinite dimensional prior over the space of functions. This implies that the estimation problem is infinite dimensional as well. However, since we sample data in a discrete manner, the GP prior becomes a multivariate Gaussian prior on $\mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_T))'$:

$$\mathbf{f} \sim \mathcal{N}(\mathbf{0}_T, K_{\vartheta}(\mathbf{X}, \mathbf{X})),$$

with $\mathbf{0}_T$ being a $T \times 1$ vector of zeros, $K_{\vartheta}(\mathbf{X}, \mathbf{X})$ a $T \times T$ kernel matrix with typical element $k_{\vartheta}(\mathbf{x}_t, \mathbf{x}_{\tau})$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)'$. This implies that, in terms of full data matrices, the GP regression is given by:

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}_T, K_{\vartheta}(\mathbf{X}, \mathbf{X})), \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T),$$

where \mathbf{I}_T denotes a $T \times T$ identity matrix.

Assuming for the moment that σ^2 is known, the posterior of \mathbf{f} follows a multivariate Gaussian distribution:

$$\mathbf{f}|\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{f}}, \bar{\mathbf{V}}_{\mathbf{f}}),$$

with variance-covariance matrix $\bar{\mathbf{V}}_{\mathbf{f}}$ and posterior mean vector $\bar{\mathbf{f}}$:

$$\begin{aligned} \bar{\mathbf{V}}_{\mathbf{f}} &= K_{\vartheta}(\mathbf{X}, \mathbf{X}) - K_{\vartheta}(\mathbf{X}, \mathbf{X}) (K_{\vartheta}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_T)^{-1} K_{\vartheta}(\mathbf{X}, \mathbf{X}), \\ \bar{\mathbf{f}} &= K_{\vartheta}(\mathbf{X}, \mathbf{X}) (K_{\vartheta}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_T)^{-1} \mathbf{y}. \end{aligned}$$

The mean function $\bar{\mathbf{f}}$ can be interpreted as a weighted average of the values of the en-

ogenous variable:

$$\bar{\mathbf{f}} = \sum_{t=1}^T \alpha_t K_{\vartheta}(\mathbf{X}, \mathbf{x}_t),$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_T)' = (K_{\vartheta}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_T)^{-1} \mathbf{y}$. This (finite dimensional) representation shows how one moves from an infinite dimensional problem to a finite dimensional one.

The expression for the variance-covariance matrix $\bar{\mathbf{V}}_{\mathbf{f}}$ also has an intuitive interpretation. The first term is the prior variance (i.e., the kernel matrix). The second term measures how much of the variance is expressed through the covariates in \mathbf{X} and thus the posterior covariance indicates how much the model learns from \mathbf{X} .

The predictive distribution of $f(\mathbf{x}_{T+h})$ can be easily derived by exploiting basic properties of the multivariate Gaussian:

$$f(\mathbf{x}_{T+h}) | \mathbf{y} \sim \mathcal{N}(\bar{f}_{T+h}, \bar{V}_{T+h}), \quad (1)$$

whereby

$$\begin{aligned} \bar{V}_{T+h} &= k_{\vartheta}(\mathbf{x}_{T+h}, \mathbf{x}_{T+h}) - K_{\vartheta}(\mathbf{x}_{T+h}, \mathbf{X}) (K_{\vartheta}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_T)^{-1} K_{\vartheta}(\mathbf{X}, \mathbf{x}_{T+h}), \\ \bar{f}_{T+h} &= K_{\vartheta}(\mathbf{x}_{T+h}, \mathbf{X}) (K_{\vartheta}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}_T)^{-1} \mathbf{y}. \end{aligned}$$

Similar to the posterior mean $\bar{\mathbf{f}}$, the predictive mean \bar{f}_{T+h} is a weighted average of the values of the endogenous variables \mathbf{y} with the weights depending on the relationship between \mathbf{X} and a realization of the vector of covariates \mathbf{x}_{T+h} related to the h -step-ahead horizon. The predictive variance \bar{V}_{T+h} , again, depends on a term that is purely driven by the prior evaluated at \mathbf{x}_{T+h} minus a term that measures the informational content in the covariates.

Before proceeding to the discussion on how to set the kernel it is worth noting that what we have discussed above is often labeled the *function-space view* of the GP. This is because the prior is elicited directly on f . Another way of analyzing GPs is based on the *weight-space view*. Under the weight-space view one can rewrite the GP regression as a

standard regression model as follows:

$$\mathbf{y} = \mathbf{W}_\vartheta \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}_T, \mathbf{I}_T),$$

with \mathbf{W}_ϑ denoting the lower Cholesky factor of $K_\vartheta(\mathbf{X}, \mathbf{X}) = \mathbf{W}_\vartheta \mathbf{W}_\vartheta'$ and $\boldsymbol{\eta}$ is a Gaussian shock vector with zero mean and unit variance. This is a standard regression model with T regressors, a coefficient vector $\boldsymbol{\eta}$ and a Gaussian prior on $\boldsymbol{\eta}$. Standard textbook formulas for the Bayesian linear regression model (see, e.g., Koop, 2003, Chapter 4) can be used to carry out posterior inference.

This also shows that if we set $K_\vartheta(\mathbf{X}, \mathbf{X}) = \mathbf{X} \mathbf{V}_\vartheta \mathbf{X}'$, we obtain a linear regression model that features a Gaussian prior with zero mean and a typical prior variance-covariance matrix \mathbf{V}_ϑ . This kernel implies many more parameters than the parsimony inducing Gaussian kernel. Hence, the resulting fit and forecasts can be expected to have posterior distributions with larger variances than those associated with the Gaussian kernel, though bias would be lower if the true model is linear and features stable parameters.

2.2 Choosing a kernel and the role of the hyperparameters

In the previous sub-section the quantities for the posterior of \mathbf{f} and the predictive density for future values of $f(\mathbf{x}_{T+h})$ suggest that the kernel and its hyperparameters play an important role. In this sub-section, we discuss this issue in more detail.

One of the key advantages of GPs is that by constructing suitable kernels, one can determine the space of possible functions. This gives rise to substantial flexibility and allows for capturing a large range of competing models within a single econometric model. For instance, Williams and Rasmussen (2006, Chapter 6) discuss how kernels can be constructed to mimic the behavior of neural networks, regression splines, polynomial and linear regressions. Tree-based techniques such as Bayesian additive regression trees (BART, Chipman, George, and McCulloch, 2010) can be cast in this framework by exploiting the ANOVA-representation of the model and then the weight-space view of the GP. In principle, and we will build on this feature later, summing over the corresponding kernels gives rise to another kernel and suitable weights could be constructed to select, in a data-driven way, which model summarizes the data best.

As stated in the previous sub-section, our focus will be on the Gaussian kernel due

to its excellent empirical properties and analytical tractability. The two hyperparameters κ and ξ control the curvature and the marginal variance of the function, respectively. We illustrate the effect of κ on the prior and posterior of \mathbf{f} in Figures 1 and 2 by means of two simple univariate examples. The first example models quarterly US inflation (in year-on-year terms) and sets $x_t = t$ for periods ranging from 2005Q1 to 2015Q4. The second example models US GDP growth as a function of the first lag of a macroeconomic uncertainty measure for the same sub-sample.⁴ The figures then show (for both the prior and posterior) the value of the function $f(x_t)$ on the y-axis and x_t on the x-axis. Both figures display in the left (right) panel the 5th and 95th prior (posterior) percentiles (with the area, the 90% credible set, in between shaded in light red) as well as three random draws from the prior (dashed red lines) in the left panel, and the posterior median (solid red lines) in the right panel.

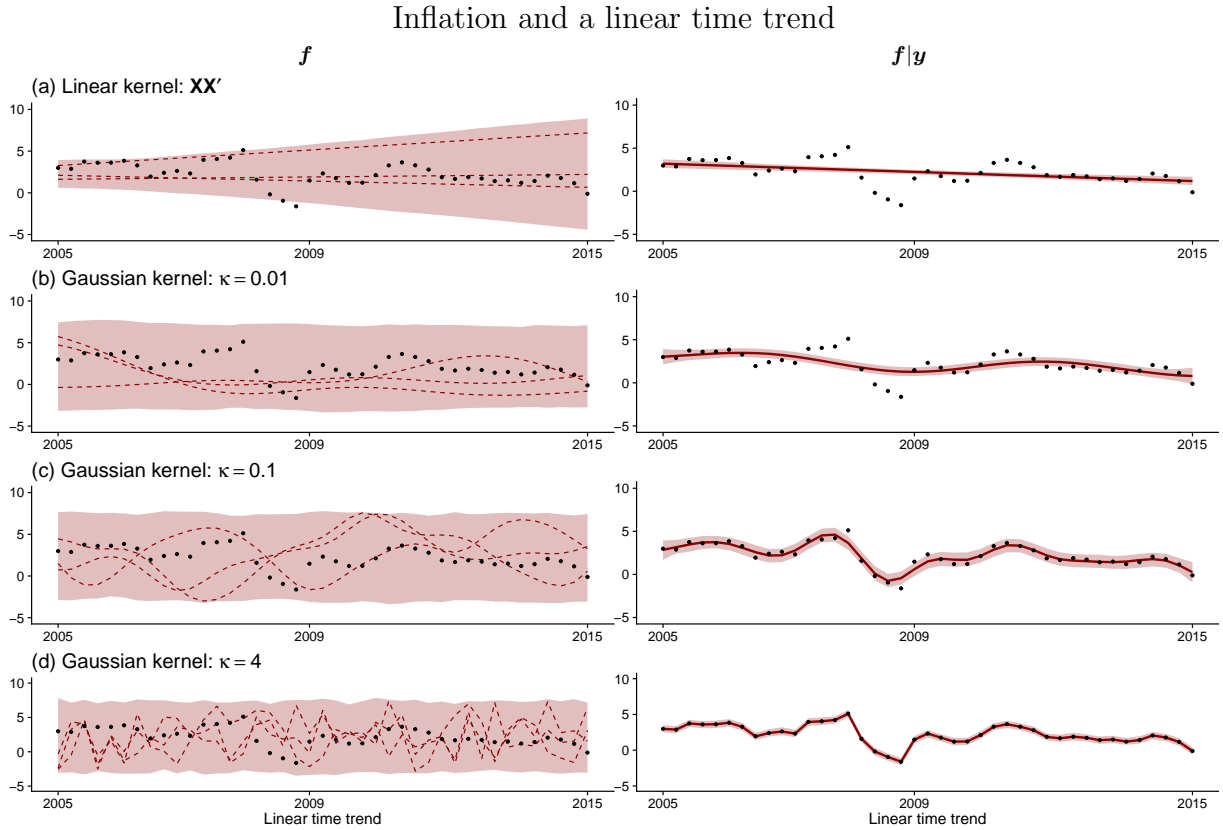
Figure 1 reveals that if y_t is a (possibly non-linear) function of time and the inverse length scale parameter is set small, the model generates functions that track the trend in inflation rather well. This is similar to the unobserved components model of Stock and Watson (2007), which features a persistent stochastic trend in inflation. Once we increase κ we observe that the draws from the prior display more high frequency variation with shorter cycles between peaks and troughs. Once this prior is combined with the data, the estimated mean functions display much more curvature and fit the actual data increasingly well.

Once κ is set too high, the functions arising from the prior vary substantially and are likely to capture also very small deviations of inflation from its trend. This translates into a close-to-perfect fit of the posterior mean of the functions and gives rise to serious overfitting concerns.

To see how a GP regression captures a possibly non-linear relationship between y_t and x_t , Figure 2 shows the functional relationship between output growth and lagged macroeconomic uncertainty. If $\kappa = 0.01$, the regression relationship is almost linear and suggests that high levels of (lagged) macroeconomic uncertainty are accompanied by negative output growth rates.

⁴Throughout the paper, we use the macroeconomic uncertainty measure of Jurado, Ludvigson, and Ng (2015) provided (and regularly updated) on the web page of Sydney C. Ludvigson (available online via sydneyludvigson.com/macro-and-financial-uncertainty-indexes). Detailed information on this index and the econometric techniques used to obtain this measure can be found in Jurado, Ludvigson, and Ng (2015).

Figure 1: Effect of different values of κ on the prior of f and the posterior $f|y$



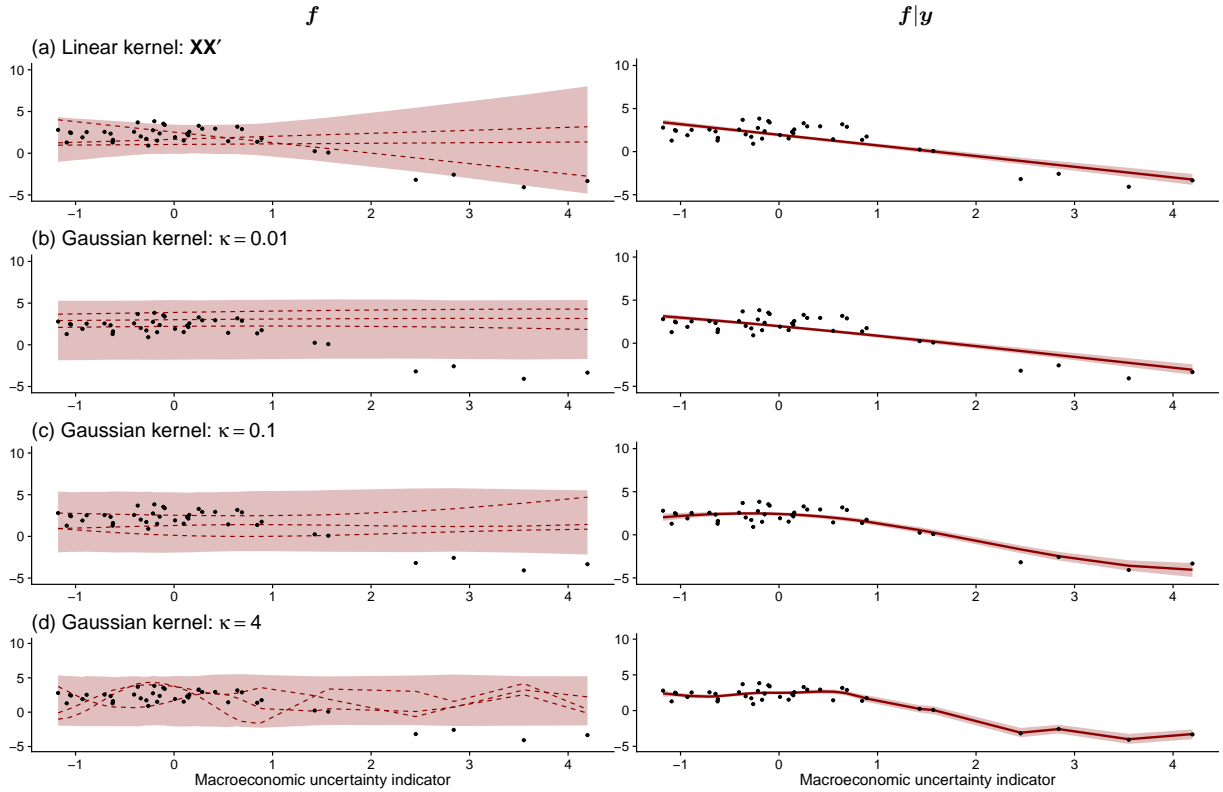
Notes: In this figure, we showcase the GP regression with US inflation data using a linear time trend as the only regressor. The left panels report, for different values of κ , the 5th and 95th prior percentiles (with the area in between shaded in light red), three draws from the prior (dashed red lines), and the actual values of inflation (black dots). The right panels report the 90% posterior credible sets (shaded in light red), the posterior medians (solid red lines), and actual inflation (black dots).

When we set $\kappa = 0.1$ we observe much more curvature (both in the prior and the posterior) in the relationship, indicating that if uncertainty is between 0 and around 1.7, GDP growth is between 2.3 and 2.5 percent. However, once a certain threshold in the first lag of uncertainty is reached, the relationship becomes strongly negative until it becomes essentially flat for very high levels of uncertainty. A similar finding, but slightly more pronounced, arises if we set $\kappa = 4$. In this case GDP growth does not change much as long as lagged uncertainty is between 0 and 1.7 and then the relationship becomes, again, strongly negative.

As is clear from these stylized examples, the role of the kernel and its hyperparameters crucially impacts the posterior estimates of the function f . Setting κ too small leads to a model which might miss important (higher frequency) information whereas a κ set too large translates into an overfitting model which might yield a very strong in-sample fit but poor out-of-sample predictions. Setting κ is thus of crucial importance and in all

Figure 2: Effect of different values of κ on the prior of f and the posterior $f|y$

GDP growth and the first lag of macroeconomic uncertainty



Notes: In this figure, we showcase the GP regression with US GDP growth data using the first lag of macroeconomic uncertainty as the only regressor. The left panels report, for different values of κ , the 5^{th} and 95^{th} prior percentiles (with the area in between shaded in light red), three draws from the prior (dashed red lines), and the actual values of GDP growth (black dots). The right panels report the 90% posterior credible sets (shaded in light red), the posterior medians (solid red lines), and actual GDP growth (black dots).

our empirical work we will infer it through Bayesian updating.

Another key question is whether the estimated function converges to the true underlying function. The literature deals with this question using several assumptions on the error distributions (mostly setting $\sigma^2 = 0$) or how the GP regression behaves if the underlying function f differs in terms of smoothness from the GP prior controlled by the kernel (see, e.g., Stone, 1982; van der Vaart and van Zanten, 2008; Yang, Bhattacharya, and Pati, 2017; Teckentrup, 2020). Stone (1982), by focusing on iid data, shows that the optimal rate of estimation of a ζ -smooth function is $T^{-\zeta/(2\zeta+K)}$ and thus decreases in K while it increases in the smoothness of the true function. Building on this finding, Teckentrup (2020) analyzes the contraction properties of a Gaussian process regression under a general Matérn kernel function and provides error bounds that also depend on the relationship of the smoothness of the true and estimated functions. If these agree, one can achieve a convergence rate of $T^{-\zeta/K}$.

After having provided the necessary foundations on Gaussian process regression, we will now focus on developing a model that is suitable for macroeconomic analysis.

3 Large-dimensional Gaussian process VARs

In this section, we first develop the GP-VAR in Sub-section 3.1. Next, Sub-sections 3.2 to 3.4 are devoted to the development of efficient MCMC schemes to carry out posterior and structural inference. Finally, Sub-section 3.5 details how to compute forecasts and (generalized) impulse response functions for the GP-VAR.

3.1 The Gaussian process VAR

In the following discussion, let $\mathbf{y}_t = (y_{1t}, \dots, y_{Mt})'$ denote an $M \times 1$ vector of macroeconomic and financial variables.⁵ Moreover, $\mathbf{x}_t = (\mathbf{x}'_{1t}, \dots, \mathbf{x}'_{Mt})'$ denotes an $Mp \times 1$ vector with $\mathbf{x}_{jt} = (y_{jt-1}, \dots, y_{jt-p})'$ storing the “own” lags of the j^{th} endogenous variable and $\mathbf{z}_t = (\mathbf{z}'_{1t}, \dots, \mathbf{z}'_{Mt})'$ an $(M-1)Mp \times 1$ vector of “other” lags. Hence, $\mathbf{z}_{jt} = (\mathbf{y}'_{-jt-1}, \dots, \mathbf{y}'_{-jt-p})'$, where \mathbf{y}_{-jt} denotes the vector \mathbf{y}_t with the j^{th} element excluded.

We discriminate between own and other lags of \mathbf{y}_t because we assume that lags of other endogenous variables impact a given endogenous variable differently from its own lags. The literature on Bayesian VARs (see Bańbura, Giannone, and Reichlin, 2010; Koop, 2013; Huber and Feldkircher, 2019; Chan, 2021) has captured this through shrinkage priors that treat coefficients on own and other lags differently. We wish to capture this equation-specific asymmetry by specifying our GP-VAR to depend on two latent processes: one driven by \mathbf{x}_t and one by \mathbf{z}_t . The structural form of the resulting GP-VAR is then given by:

$$\mathbf{y}_t = F(\mathbf{x}_t) + G(\mathbf{z}_t) + \mathbf{Q}\mathbf{y}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \mathbf{H}_t), \quad (2)$$

with $F(\mathbf{x}_t) = (f_1(\mathbf{x}_{1t}), \dots, f_M(\mathbf{x}_{Mt}))'$ and $G(\mathbf{z}_t) = (g_1(\mathbf{z}_{1t}), \dots, g_M(\mathbf{z}_{Mt}))'$ and f_j and g_j being equation-specific functions. The function f_j controls how y_{jt} depends on its own lags while g_j encodes the relationship between y_{jt} and the lags of the other endogenous variables. The functions f_j and g_j , and hence F and G , differ because we construct differ-

⁵We assume that the elements in \mathbf{y}_t are demeaned. In our empirical application we include a constant term with an uninformative prior.

ent kernels with distinct hyperparameters.⁶ The matrix \mathbf{Q} is an $M \times M$ lower triangular matrix with zeros along its main diagonal. This matrix defines the contemporaneous relations across the elements in \mathbf{y}_t .

Finally, $\boldsymbol{\varepsilon}_t$ is an $M \times 1$ vector of Gaussian shocks with zero mean and an $M \times M$ time-varying variance-covariance matrix $\mathbf{H}_t = \text{diag}(\omega_{1t}, \dots, \omega_{Mt})$. We will assume that ω_{jt} follows a flexible stochastic volatility (SV) model:

$$h_{jt} = \log \omega_{jt} = \rho_{hj} h_{jt-1} + \nu_{h,jt}, \quad \nu_{h,jt} \sim \mathcal{N}(0, \sigma_{hj}^2), \quad h_{j0} \sim \mathcal{N}\left(0, \frac{\sigma_{hj}^2}{1 - \rho_{hj}^2}\right), \quad (3)$$

with the logarithm of $h_{jt} = \log \omega_{jt}$ being assumed to evolve according to a stationary AR(1) state equation. We let ρ_{hj} denote the persistence parameter, σ_{hj}^2 the error variance, and h_{j0} the initial state of the log-volatility process.

Allowing for time variation in the shock variances provides additional flexibility and enables us to capture non-Gaussian features in the shocks (not only, but also due to the fact that h_{jt} enters the model non-linearly).⁷ In principle, we could also allow for unknown functional relations between the contemporaneous terms of the preceding $j - 1$ equations and the response of equation j . However, this would lead to a complicated non-linear covariance structure. Since we are interested in carrying out structural identification based on zero impact restrictions we opt for choosing this simpler approach which implies multivariate Gaussian reduced form shocks, but with a time-varying covariance matrix. Given that the literature on GPs typically assumes the shocks to be Gaussian and homoskedastic, this is already a substantial increase in flexibility.⁸

The model in Eq. (2) assumes that the shocks in $\boldsymbol{\varepsilon}_t$ are, conditional on $\mathbf{Q}\mathbf{y}_t$, orthogonal and hence estimation can be carried out equation-by-equation. We will exploit this representation for simplicity and computational tractability. The j^{th} equation, in terms of full-data matrices, is given by:

$$\mathbf{Y}_j = \mathbf{f}_j + \mathbf{g}_j + \sum_{k=1}^{j-1} q_{jk} \mathbf{Y}_k + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{\varepsilon}_j \sim \mathcal{N}(\mathbf{0}_T, \boldsymbol{\Omega}_j),$$

⁶It is worth stressing that one could also think of our decomposition in terms of a new function with a kernel that is given by the sum of the kernels of the functions f_j and g_j .

⁷One could also introduce additional scaling factors that arise from inverse Gamma distributions to obtain a model with t -distributed shocks.

⁸A rare exception is Jylänki, Vanhatalo, and Vehtari (2011), who propose a GP regression with heavy tailed errors and mainly focus on fast and robust approximate inference of a posterior that is analytically intractable due to a t -distributed likelihood.

with $\mathbf{Y}_j = (y_{j1}, \dots, y_{jT})'$, $\mathbf{f}_j = (f_j(\mathbf{x}_{j1}), \dots, f_j(\mathbf{x}_{jT}))'$, $\mathbf{g}_j = (g_j(\mathbf{z}_{j1}), \dots, g_j(\mathbf{z}_{jT}))'$, $\mathbf{\Omega}_j = \text{diag}(\omega_{j1}, \dots, \omega_{jT})$, $\boldsymbol{\epsilon}_j = (\varepsilon_{j1}, \dots, \varepsilon_{jT})'$ and q_{jk} denoting the $(j, k)^{\text{th}}$ element of \mathbf{Q} . We will use this form to carry out inference about the unknown functions f_j and g_j as well as the remaining parameters and latent states of the model.

Notice that our estimation strategy is not invariant with respect to reordering the elements in \mathbf{y}_t , a common problem if this orthogonalization strategy is used. In Sub-section C.2 of the Online Appendix, we show that different orderings have only a small impact on the estimated impulse responses.

3.2 Conjugate Gaussian process priors

In this sub-section, our focus will be on the priors on f_j and g_j . The priors on the remaining, linear quantities are standard and thus not discussed in depth. We use a Horseshoe prior (Carvalho, Polson, and Scott, 2010) on the free elements in \mathbf{Q} , a Beta prior on the (transformed) persistence parameter $(\rho_{hj} + 1)/2 \sim \mathcal{B}(25, 5)$, and an inverse Gamma prior on the state innovation variances σ_{hj}^2 . This prior is specified to have mean 0.1 and variance 0.01.

For equation-specific functions f_j and g_j , we specify two GPs with one conditional on $\mathbf{X}_j = (\mathbf{x}_{j1}, \dots, \mathbf{x}_{jT})'$ and one conditional on $\mathbf{Z}_j = (\mathbf{z}_{j1}, \dots, \mathbf{z}_{jT})'$:

$$\mathbf{f}_j \sim \mathcal{N}\left(\mathbf{0}_T, \sqrt{\mathbf{\Omega}_j} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \sqrt{\mathbf{\Omega}_j}\right), \quad \mathbf{g}_j \sim \mathcal{N}\left(\mathbf{0}_T, \sqrt{\mathbf{\Omega}_j} K_{\boldsymbol{\vartheta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) \sqrt{\mathbf{\Omega}_j}\right). \quad (4)$$

We let $\sqrt{\mathbf{\Omega}_j} = \text{diag}(\sqrt{\omega_{j1}}, \dots, \sqrt{\omega_{jT}})$ while $K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j)$ and $K_{\boldsymbol{\vartheta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j)$ denote two suitable kernels with typical elements given by:

$$\begin{aligned} k_{\boldsymbol{\vartheta}_{j1}}(\mathbf{x}_{jt}, \mathbf{x}_{j\tau}) &= \xi_{j1} \times \exp\left(-\frac{\kappa_{j1}}{2}(\mathbf{x}_{jt} - \mathbf{x}_{j\tau})' \mathbf{D}_{\mathbf{X}_j}^{-1}(\mathbf{x}_{jt} - \mathbf{x}_{j\tau})\right), \quad \boldsymbol{\vartheta}_{j1} = (\xi_{j1}, \kappa_{j1})', \\ k_{\boldsymbol{\vartheta}_{j2}}(\mathbf{z}_{jt}, \mathbf{z}_{j\tau}) &= \xi_{j2} \times \exp\left(-\frac{\kappa_{j2}}{2}(\mathbf{z}_{jt} - \mathbf{z}_{j\tau})' \mathbf{D}_{\mathbf{Z}_j}^{-1}(\mathbf{z}_{jt} - \mathbf{z}_{j\tau})\right), \quad \boldsymbol{\vartheta}_{j2} = (\xi_{j2}, \kappa_{j2})'. \end{aligned}$$

For $j = 1, \dots, M$, $\boldsymbol{\vartheta}_{j1}$ and $\boldsymbol{\vartheta}_{j2}$ are equation and kernel-specific hyperparameters and the matrices $\mathbf{D}_{\mathbf{X}_j}, \mathbf{D}_{\mathbf{Z}_j}$ are diagonal matrices with typical i^{th} element $\hat{\sigma}_{\mathbf{X}_j i}^2, \hat{\sigma}_{\mathbf{Z}_j i}^2$. These are set equal to the empirical variances of the i^{th} column of \mathbf{X}_j and \mathbf{Z}_j , respectively. Inclusion of the diagonal scaling matrices $\mathbf{D}_{\mathbf{X}_j}$ and $\mathbf{D}_{\mathbf{Z}_j}$ serves to control for differences in the scaling of the explanatory variables. Notice that since the hyperparameters are allowed to differ, we essentially treat own and other lags asymmetrically through different

functional approximations f_j and g_j .

The kernel is scaled with the error variances in $\mathbf{\Omega}_j$. A typical diagonal element of the corresponding re-scaled kernel is given by $\omega_{jt} \times k_{\boldsymbol{\theta}_{j1}}(\mathbf{x}_{jt}, \mathbf{x}_{jt}) = \omega_{jt}\xi_{j1}$ and $\omega_{jt} \times k_{\boldsymbol{\theta}_{j2}}(\mathbf{z}_{jt}, \mathbf{z}_{jt}) = \omega_{jt}\xi_{j2}$. Typical off-diagonal elements are given by $\sqrt{\omega_{jt}}\sqrt{\omega_{j\tau}} \times k_{\boldsymbol{\theta}_{j1}}(\mathbf{x}_{jt}, \mathbf{x}_{j\tau})$ and $\sqrt{\omega_{jt}}\sqrt{\omega_{j\tau}} \times k_{\boldsymbol{\theta}_{j2}}(\mathbf{z}_{jt}, \mathbf{z}_{j\tau})$. The interaction between the kernel and the error variances gives rise to convenient statistical and computational properties.

First, note that if ω_{jt} is large, the corresponding prior on the unknown functions is more spread out. In macroeconomic data, ω_{jt} is typically large in crisis periods when the \mathbf{x}_{jt} and \mathbf{z}_{jt} are far away from their previous values. Since the diagonal elements of the kernels are effectively determined by $\boldsymbol{\xi}_j = (\xi_{j1}, \xi_{j2})'$ the presence of ω_{jt} allows for larger values in the marginal prior variance and thus makes large shifts in the unknown functions more likely. Second, the interaction between ω_{jt} and ω_{jt-1} implies that the covariances are scaled down if $\omega_{jt} \gg \omega_{jt-1}$, suggesting that the informational content decreases if increases in uncertainty are substantial (i.e., $\Delta\omega_{jt}$ is large). If $\omega_{jt} \approx \omega_{j\tau}$ and both are large, the corresponding covariance will be scaled upwards. This implies that our model learns from previous crisis episodes as well. Third, as we will show in Sub-section 3.4, interacting the kernel with the error variances leads to a conjugate Gaussian process structure which implies that we can factor out the error volatilities and do not need to update several quantities during MCMC sampling. This speeds up computation enormously and allows for estimating large models.

Before discussing how we select the hyperparameters, it is worth highlighting a possible identification problem of our model. In our baseline specification we center \mathbf{f}_j and \mathbf{g}_j around zero a priori. If we introduce an additional intercept term (or a simpler mean function) no identification issues arise. However, if we believe that \mathbf{f}_j and \mathbf{g}_j are centered on non-zero values, we can not separately identify them. In our empirical work, we normalize the grand mean of \mathbf{g}_j to be equal to zero.⁹

It is worth stressing, however, that if interest is on predictions or impulse responses, this does not cause any additional issues since the conditional mean function (which is the sum over \mathbf{f}_j and \mathbf{g}_j) is identified. Exploiting basic properties of the Gaussian distribution one can easily show that the sum of \mathbf{f}_j and \mathbf{g}_j in Eq. (4) gives rise to a new latent process

⁹Notice that this only concerns the posterior distribution since, under the prior, this condition is automatically fulfilled.

\mathbf{m}_j which is, again, Gaussian:

$$\mathbf{m}_j \sim \mathcal{N}\left(\mathbf{0}_T, \sqrt{\Omega_j} (K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + K_{\boldsymbol{\vartheta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j)) \sqrt{\Omega_j}\right).$$

Hence, our model can be also viewed as a standard Gaussian process that combines information in \mathbf{X}_j and \mathbf{Z}_j by summing over two different kernels. This immediately implies that if we are interested in sampling from the posterior predictive distribution of \mathbf{y}_{t+h} (and related functions such as impulse responses) it is sufficient to estimate \mathbf{m}_j .

3.3 Selecting the hyperparameters associated with the kernel

So far, we always conditioned on the hyperparameters that determine the shape of the Gaussian kernel. A simple way of specifying $\boldsymbol{\vartheta}_{j1}$ and $\boldsymbol{\vartheta}_{j2}$ is the *median heuristic* approach stipulated in Chaudhuri, et al. (2017). This choice works well in a wide range of applications featuring many covariates (see, e.g., Crawford, et al., 2019). The median heuristic fixes $\xi_{j1} = \xi_{j2} = 1$ and defines the inverse of the bandwidth parameter as:

$$\kappa_{j1} = \bar{\kappa}_{j1} = \text{median}_{t\tau} \left(\frac{1}{\|\mathbf{x}_{jt} - \mathbf{x}_{j\tau}\|} \right), \quad \kappa_{j2} = \bar{\kappa}_{j2} = \text{median}_{t\tau} \left(\frac{1}{\|\mathbf{z}_{jt} - \mathbf{z}_{j\tau}\|} \right),$$

for $j = 1, \dots, M$. This simple approach has the convenient property that it automatically selects a bandwidth which is consistent with the time series behavior of the elements in \mathbf{y}_t . To illustrate this, suppose that y_{jt} is a highly persistent process (e.g., inflation or short-term interest rates). In this case, for $\tau = t - 1$, the Euclidean distance $\|\mathbf{x}_{jt} - \mathbf{x}_{j\tau}\|$ will be quite small and, hence, the mean function \mathbf{f}_j smoothly adjusts. If y_{jt} is less persistent and displays large fluctuations (e.g., stock market or exchange rate returns), the Euclidean distance $\|\mathbf{x}_{jt} - \mathbf{x}_{j\tau}\|$ will be large and, thus, \mathbf{f}_j allows for capturing this behavior. The dispersion in \mathbf{z}_{jt} might have important implications for y_{jt} if the aim is to model a trend in y_{jt} that depends on other covariates. This could arise in a situation where the prior on \mathbf{f}_j is set very tight (i.e., the posterior of \mathbf{f}_j will be centered on zero) and information not coming from \mathbf{x}_{jt} would then determine the behavior of y_{jt} . This discussion highlights how the median heuristic allows for flexibly discriminating between signal and noise and thus acts as a non-linear filter which purges the time series from high frequency variation, if necessary.

Given that we work with potentially large panels of time series, it is questionable

that the median heuristic works equally well for all elements in \mathbf{y}_t . As a solution, we propose to use the median heuristic to set up a discrete grid for both ξ_{j1} (ξ_{j2}) and κ_{j1} (κ_{j2}). For each element in this grid we specify a hyperprior. We use Gamma priors on all elements. For $j = 1, \dots, M$, that is

$$\xi_{j1} \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2c_{\xi 1}}\right) \quad \text{and} \quad \xi_{j2} \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2c_{\xi 2}}\right),$$

for the linear shrinkage hyperparameters and

$$\kappa_{j1} \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2c_{\kappa 1}}\right) \quad \text{and} \quad \kappa_{j2} \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2c_{\kappa 2}}\right),$$

for the bandwidth parameters. Here, $c_{\xi 1}, c_{\xi 2}, c_{\kappa 1}$ and $c_{\kappa 2}$ are scalars that define the tightness of the hyperprior. In the empirical application, we set $c_{\xi 1} = c_{\xi 2} = c_{\xi}$ and $c_{\kappa 1} = c_{\kappa 2} = c_{\kappa}$. These parameters strongly influence the shape of the conditional mean and are crucial modeling choices and we set them through cross-validation. In our empirical application, we find that small values of c_{κ} work reasonably well, yielding an informative prior that forces κ_{j1} and κ_{j2} towards zero.

Based on this set of priors we can derive the conditional posterior distribution. Since ξ_{j1} (ξ_{j2}) and κ_{j1} (κ_{j2}) are placed on a grid, we can pre-compute several quantities related to the kernel (such as inverses and Cholesky factors) while at the same time infer them from the data with sufficient accuracy, which is crucial for precise inference. In what follows, we center these grids around the median heuristic and additionally take into account the considerations of the informative Gamma priors:

$$\kappa_{jk} \in [0.1\bar{\kappa}_{jk}, 2\bar{\kappa}_{jk}] \quad \text{and} \quad \xi_{jk} \in [0.04, 4], \quad \text{for } j = 1, \dots, M \text{ and } k = 1, 2.$$

Here, the intervals indicate the minimum (maximum) value supported for each hyperparameter. Within this two dimensional range, we define a discrete grid of around 1000 combinations with equally sized increments along each dimension.¹⁰ The corresponding posterior is discrete and we can use inverse transform sampling to carry out posterior inference. Further details are provided in Sub-section 3.4.

¹⁰Implicitly, this two dimensional grid results in a prior view in which any hyperparameter combination not included in the grid has zero support.

3.4 Posterior computation

Posterior inference for the GP-VAR is carried out using a novel yet conceptually simple MCMC algorithm which cycles between several steps. In this section we will focus on how to sample from the posterior of \mathbf{f}_j , $p(\mathbf{f}_j|\bullet)$, with \bullet denoting conditioning on everything else, and $\boldsymbol{\vartheta}_{j1}$. Sampling from $p(\mathbf{g}_j|\bullet)$ and $p(\boldsymbol{\vartheta}_{j2}|\bullet)$ works analogously with some adjustments. These relate to the fact that we introduce a linear restriction that $(\boldsymbol{\iota}'\boldsymbol{\iota})^{-1}\boldsymbol{\iota}'\mathbf{g}_j = 0$, with $\boldsymbol{\iota}$ denoting a $T \times 1$ vector of ones. The corresponding conditional posterior distribution is a hyperplane truncated Gaussian where efficient sampling algorithms are available (see Cong, Chen, and Zhou, 2017). Further details can be found in Sub-section A.2 of the Online Appendix. It is worth stressing that we sample \mathbf{f}_j and \mathbf{g}_j separately. This increases the computational burden slightly but allows us to consider both latent processes separately from each other. In case our focus is purely on prediction or impulse response analysis, one can also simulate the process \mathbf{m}_j defined in Sub-section 3.2 without any additional restriction. Both procedures yield exactly the same results.

Generalizing the results in Section 2, it can be shown that the posterior of \mathbf{f}_j is Gaussian for all j :

$$\mathbf{f}_j|\bullet \sim \mathcal{N}(\bar{\mathbf{f}}_j, \bar{\mathbf{V}}_{\mathbf{f}_j}),$$

with posterior moments given by:

$$\begin{aligned} \bar{\mathbf{V}}_{\mathbf{f}_j} &= \sqrt{\Omega_j} \left(K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) - K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) (K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + \mathbf{I}_T)^{-1} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \right) \sqrt{\Omega_j}, \\ \bar{\mathbf{f}}_j &= \sqrt{\Omega_j} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) (K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + \mathbf{I}_T)^{-1} \sqrt{\Omega_j}^{-1} \left(\mathbf{Y}_j - \mathbf{g}_j - \sum_{k=1}^{j-1} q_{jk} \mathbf{Y}_k \right), \end{aligned}$$

where for $j = 1$, the term $\sum_{k=1}^{j-1} q_{jk} \mathbf{Y}_k$ is excluded.

In principle, computing the inverse and the Cholesky factor of $\bar{\mathbf{V}}_{\mathbf{f}_j}^{-1}$ constitutes the main bottleneck when it comes to sampling from $p(\mathbf{f}_j|\bullet)$. This is especially so if T is large. But, in common macroeconomic applications which use quarterly US data, T is moderate and thus computation is feasible. In our case, even if interest centers on using monthly data or even higher frequencies, we can exploit the convenient fact that, conditional on

the hyperparameters ξ_{j1} and κ_{j1} ,

$$\mathbf{B}_{f_j} \mathbf{B}'_{f_j} = \left(K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) - K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \left(K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + \mathbf{I}_T \right)^{-1} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \right)$$

as well as its Cholesky factor \mathbf{B}_{f_j} can be pre-computed. In addition, notice that

$$\bar{\mathbf{V}}_{f_j} = \mathbf{C}_{f_j} \mathbf{C}'_{f_j} = \left(\sqrt{\boldsymbol{\Omega}_j} \mathbf{B}_{f_j} \right) \left(\sqrt{\boldsymbol{\Omega}_j} \mathbf{B}_{f_j} \right)'.$$

These practical properties (due to the conjugate structure) substantially speed up computation in terms of sampling from $p(\mathbf{f}_j | \bullet)$.

These results are conditional on the hyperparameters. As outlined in the previous sub-section, we will estimate them by defining a discrete two dimensional grid of 1000 combinations. For each hyperparameter combination on this grid, we compute the corresponding kernel $K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j)$ as well as all relevant quantities (i.e., \mathbf{C}_{f_j}). Based on these values we jointly evaluate the conditional posterior ordinate by applying Bayes theorem. The exact form of the conditional likelihood is given by:

$$p(\mathbf{f}_j | \boldsymbol{\vartheta}_{j1}, \boldsymbol{\Omega}_j) = (2\pi)^{-\frac{T}{2}} \times \det \left(\sqrt{\boldsymbol{\Omega}_j} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \sqrt{\boldsymbol{\Omega}_j} \right)^{-\frac{1}{2}} \\ \times \exp \left\{ -\frac{1}{2} \left(\mathbf{f}'_j \left(\sqrt{\boldsymbol{\Omega}_j} K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) \sqrt{\boldsymbol{\Omega}_j} \right)^{-1} \mathbf{f}_j \right) \right\}.$$

Note that the shape of $K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j)$ depends on the hyperparameters $\boldsymbol{\vartheta}_{j1} = (\xi_{j1}, \kappa_{j1})'$, which we want to update. For each pair of values $\boldsymbol{\vartheta}_{j1}^{(s)}$ on our two dimensional grid (with s denoting a specific combination), we compute the corresponding kernel $K_{\boldsymbol{\vartheta}_{j1}=\boldsymbol{\vartheta}_{j1}^{(s)}}(\mathbf{X}_j, \mathbf{X}_j)$ as well as $\det \left(K_{\boldsymbol{\vartheta}_{j1}=\boldsymbol{\vartheta}_{j1}^{(s)}}(\mathbf{X}_j, \mathbf{X}_j) \right)$ and $\left(K_{\boldsymbol{\vartheta}_{j1}=\boldsymbol{\vartheta}_{j1}^{(s)}}(\mathbf{X}_j, \mathbf{X}_j) \right)^{-1}$ prior to MCMC sampling. Hence, within our sampler evaluating the likelihood is straightforward and computationally efficient. All that remains is to multiply the likelihood with the prior. The corresponding posterior ordinates for each $\boldsymbol{\vartheta}_{j1}^{(s)}$ are used to compute probabilities to perform inverse transform sampling to sample from $p(\boldsymbol{\vartheta}_{j1} | \bullet)$.

Conditional on \mathbf{f}_j and \mathbf{g}_j , the remaining parameters (i.e., the free elements in \mathbf{Q} , the log-volatilities and the associated coefficients in the corresponding state equations) can be sampled through (mostly) standard steps. One modification relates to how we sample the volatilities in $\boldsymbol{\Omega}_j$. The main difference stems from the fact that the volatilities in $\boldsymbol{\Omega}_j$ also show up in the prior on \mathbf{f}_j and \mathbf{g}_j . To circumvent this issue we integrate out the

latent processes \mathbf{f}_j and \mathbf{g}_j . This calls for a minor adjustment of the original sampler by integrating out the latent processes \mathbf{f}_j and \mathbf{g}_j first and then sampling the log-volatilities using an independent Metropolis Hastings update similar to the one proposed in Chan (2017). We provide additional details and the full posterior simulator in Section C of the Online Appendix.

3.5 Forecasts and generalized impulse responses

In non-parametric models such as the GP regression described in Section 2, the effect of the covariates on \mathbf{y}_t are typically analyzed through so-called partial dependence plots (Friedman, 2001). Our large dimensional setting and the fact that we have a VAR-type structure in the conditional mean, imply that partial dependence plots are difficult to compute and visualize since they would require integration over a large number of covariates. Moreover, VARs are dynamic models and partial dependence plots are difficult to employ in dynamic settings. To capture non-linear model dynamics and possible relations across variables, impulse responses are used to investigate the effects of structural shocks on \mathbf{y}_t . In this paper, we will follow this route as well. Because the model is highly non-linear, we need to resort to generalized impulse responses (GIRFs) originally proposed in Koop, Pesaran, and Potter (1996). As GIRFs are based on (multi-step-ahead) forecasts, and since forecasting with the GP-VAR is of interest by itself, we start with a discussion on forecast computation.

We begin by computing the predictive distribution of the one-step-ahead forecasts $p(\mathbf{y}_{t+1}|\mathcal{I}_t)$, with \mathcal{I}_t denoting all available information up to time t . The one-step-ahead predictive distribution is obtained by simulating from the predictive distribution of \mathbf{m}_{t+1} , $p(\mathbf{m}_{t+1}|\mathcal{I}_t)$, and sampling from the marginal distribution of the shocks $\boldsymbol{\varepsilon}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_{t+1})$. The draw from the one-step-ahead predictive density is used to set up \mathbf{x}_{t+2} and \mathbf{z}_{t+2} . Based on these, we can compute the corresponding kernels and obtain a draw \mathbf{m}_{t+2} from the density $p(\mathbf{m}_{t+2}|\mathcal{I}_t)$. Again, a draw from $\mathbf{y}_{t+2} \sim p(\mathbf{y}_{t+2}|\mathcal{I}_t)$ is obtained by sampling from the marginal shock distribution $\boldsymbol{\varepsilon}_{t+2} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_{t+2})$ and adding this draw to \mathbf{m}_{t+2} . Higher order forecasts are obtained analogously. The resulting predictive distribution of \mathbf{y}_{t+h} will be highly non-Gaussian and might feature heavy tails and/or asymmetries. This forms the baseline.

The GIRFs are computed as follows. To analyze the effects of a structural distur-

bance (such as an uncertainty shock) we assume that the uncertainty indicator is (without loss of generality) in the j^{th} position in \mathbf{y}_t . A corresponding shock of size ς in time t to the uncertainty indicator shifts all elements in \mathbf{y}_t by $\varsigma \mathbf{q}_j$, i.e., the j^{th} column of $(\mathbf{I} - \mathbf{Q})^{-1}$ scaled by a scalar that reflects the shock size ς . The other shocks are sampled, again, from their marginal distributions, i.e., for all $i \neq j$ we have that $\varepsilon_{it} \sim \mathcal{N}(0, \omega_{it})$. Based on this we draw from the predictive distribution conditional on the uncertainty shock $\hat{\mathbf{y}}_{t+1} \sim p(\mathbf{y}_{t+1} | \mathcal{I}_t, \varepsilon_{jt} = \varsigma)$ and use this draw to compute $\hat{\mathbf{x}}_{t+2}$ and $\hat{\mathbf{z}}_{t+2}$. For higher order conditional forecasts we proceed as in the case of the unconditional forecast distribution by simulating from the marginal distribution of the structural shocks which are added to the conditional mean forecasts $\hat{\mathbf{m}}_{t+h}$.

The corresponding dynamic responses are then obtained by subtracting the mean of the unconditional predictive distribution from the conditional (on the uncertainty shock) predictive density. This yields:

$$\delta_{ht} = \mathbb{E}(\mathbf{y}_{t+h} | \mathcal{I}_t, \varepsilon_{jt} = \varsigma) - \mathbb{E}(\mathbf{y}_{t+h} | \mathcal{I}_t).$$

Notice that δ_{ht} is state-dependent and, due to the non-linear nature of the conditional mean function, allows for asymmetries in how \mathbf{y}_t reacts to shocks.¹¹ This gives rise to two inferential opportunities. First, one can assess how a given shock has impacted the economy in a given point in time. This allows us to investigate whether transmission mechanisms depend on the underlying state of the economy. Second, the non-linear mean function directly implies that shock transmission can be asymmetric, so that positive shocks might feed through the economy differently than negative shocks, and non-proportional, so that larger shocks can have proportionally different effects than smaller shocks. In all our empirical work we will exploit both dimensions and focus on asymmetries in the sign, and non-proportionality in the size, of the shock as well as explicitly consider state dependencies by computing δ_{ht} over time. Finally, we also integrate out uncertainty with respect to the state of the economy by averaging over all values of t .

¹¹The full details on computing generalized impulse response functions can be found in Section A.5 of the Online Appendix.

4 Illustration using synthetic data

In this section, we illustrate the computational merits of our approach and evaluate whether it successfully recovers different features of a highly non-linear DGP.

To illustrate our methods, we simulate $T = 200$ observations from a highly non-linear small-scale VAR with $M = 3$ equations. The three equations differ in terms of whether they are linear or non-linear in the parameters but also with respect to the distribution of the shocks. Non-linearities are captured in two ways. First, we assume a break point and second we assume non-linear relations between the response variable and the lags of the other variables. In all these equations, we assume that the functions f_j and g_j differ to assess whether our approach is capable of discriminating between the two. The precise form of our DGP is given by:

$$\mathbf{y}_t = F(\mathbf{x}_t) + G(\mathbf{z}_t) + \mathbf{Q}\mathbf{y}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \mathbf{H}_t),$$

with

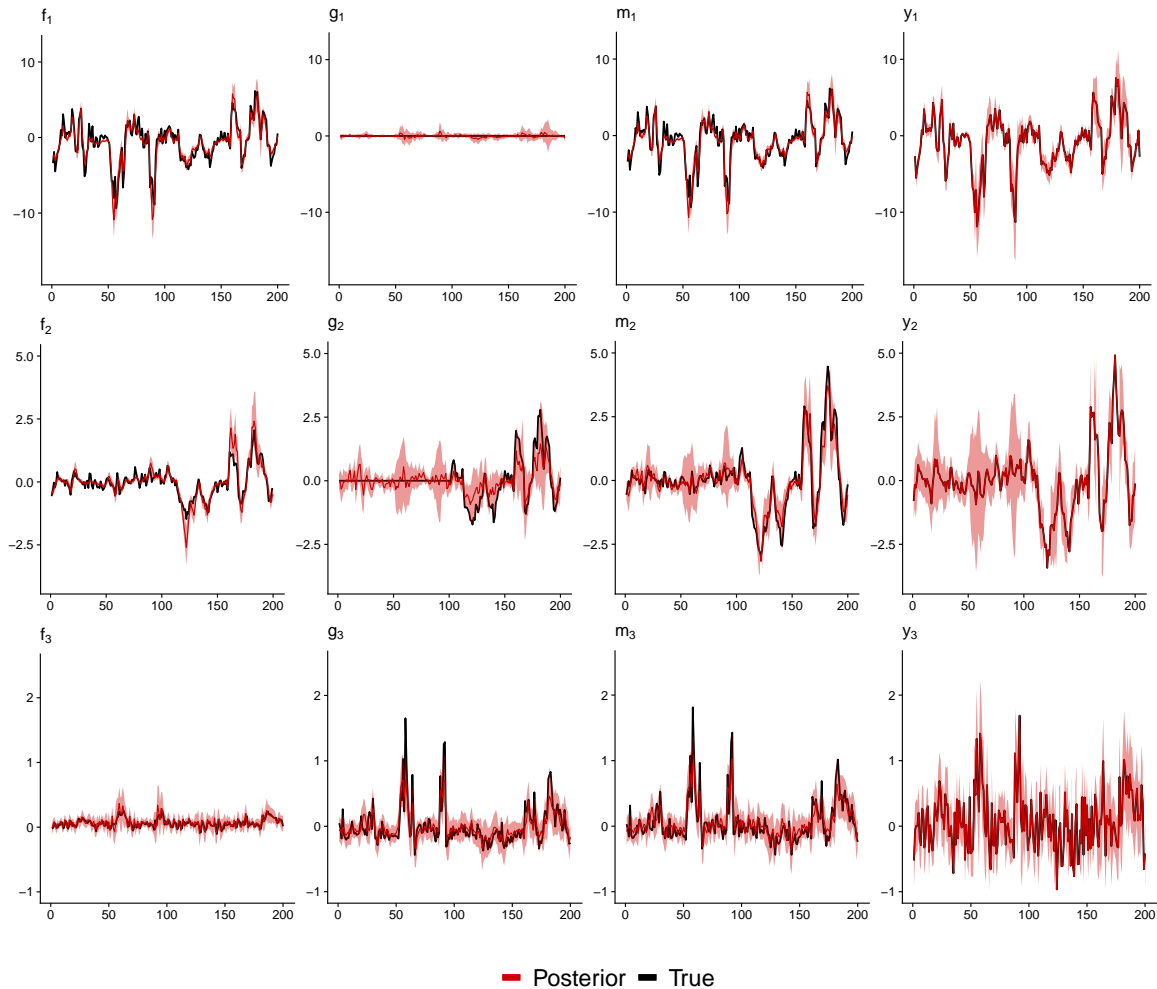
$$F(\mathbf{x}_t) = \begin{pmatrix} f_1(\mathbf{x}_{1t}) \\ f_2(\mathbf{x}_{2t}) \\ f_3(\mathbf{x}_{3t}) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^p \phi_{11,k} y_{1t-k} \\ \sum_{k=1}^p \phi_{22,k} y_{2t-k} \times \mathcal{I}(t \leq 100) + \phi_{22,1} y_{2t-1} \times \mathcal{I}(t > 100) \\ \frac{1}{12} \sin(\frac{\pi}{2} y_{3t-1} y_{3t-2}) + \frac{1}{3} (y_{3t-3} - 1)^2 + \frac{1}{12} y_{3t-4} + \frac{1}{12} y_{3t-5} \end{pmatrix},$$

$$G(\mathbf{z}_t) = \begin{pmatrix} g_1(\mathbf{z}_{1t}) \\ g_2(\mathbf{z}_{2t}) \\ g_3(\mathbf{z}_{3t}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \times \mathcal{I}(t \leq 100) + \sum_{k=1}^p \sum_{j \in \{1,3\}} \phi_{2j,k} y_{jt-k} \times \mathcal{I}(t > 100) \\ \frac{1}{18} \sin(\frac{\pi}{2} y_{1t-1} y_{2t-1}) + \frac{2}{9} (y_{1t-2} - 1)^2 + \frac{1}{18} y_{1t-3} + \frac{1}{18} y_{2t-5} \end{pmatrix},$$

where $\mathbf{H}_t = \text{diag}(\omega_{1t}, \omega_{2t}, \omega_{3t})$, $\phi_{11,1} = 0.8$, $\phi_{22,1} = 0.65$, $\phi_{ij,k} \sim \mathcal{N}\left(0, \left(\frac{0.3}{k}\right)^2\right)$ for $i \neq j$ and $k = 1, \dots, p$, and $\mathcal{I}(\bullet)$ denotes the indicator function that equals one if its argument is true and zero otherwise. The free elements in \mathbf{Q} , are also simulated from a Gaussian distribution with $q_{jk} \sim \mathcal{N}(0, 0.1^2)$. Moreover, we introduce an SV specification with heavy tails for the structural error variances $\omega_{jt} = \lambda_{jt} \tilde{\omega}_{jt}$ with each $\tilde{h}_{jt} = \log \tilde{\omega}_{jt}$ following an independent random walk law of motion: $\tilde{h}_{jt} = \tilde{h}_{jt-1} + \sigma_{\tilde{h}_j} u_{\tilde{h}_t}$, with $u_{\tilde{h}_t} \sim \mathcal{N}(0, 1)$. For each equation, we set the initial state $\tilde{\omega}_{j0} = \exp \tilde{h}_{j0} = 0.01$ and the state innovation variance $\sigma_{\tilde{h}_j} = 0.01$. We consider (conditionally) t_3 -distributed errors with three degrees of freedom in the first equation by simulating $\lambda_{1t} \sim \mathcal{G}^{-1}(3/2, 3/2)$ and (conditionally) Gaussian-distributed errors for the second and third equation by setting $\lambda_{2t} = \lambda_{3t} = 1$ for

all t .

Figure 3: Posterior distributions of \mathbf{f}_j , \mathbf{g}_j , $\mathbf{m}_j(= \mathbf{f}_j + \mathbf{g}_j)$ and \mathbf{y}_j versus actual realizations for each of the three equations in the DGP.



Notes: Results are obtained from simulating a single realization from the proposed DGP. Horizontal panels refer to each equation of the DGP. The solid black lines denote the actual outcomes of the respective functions, while the red solid lines represent the posterior medians and the red shaded areas the 90% posterior credible sets of the respective fitted values.

Figure 3 shows results obtained from simulating a single realization from the DGP. Horizontal panels refer to each equation of the DGP while vertical panels show the different components. The red shaded areas represent the 90% posterior credible set of \mathbf{f}_j , \mathbf{g}_j , $\mathbf{m}_j(= \mathbf{f}_j + \mathbf{g}_j)$ and \mathbf{y}_j , while the solid black line denotes the actual outcome of these quantities. The figure suggests that our approach is capable of detecting different functional relations between \mathbf{y}_t , \mathbf{x}_t and \mathbf{z}_t . Across all three equations, we find that the estimated conditional mean function \mathbf{m}_j tracks the actual value rather well. It is also worth stressing that if the DGP features non-Gaussian shocks, the model recovers the true mean function particularly well (see the first row of Figure 3). Zooming into the estimates for the different

Table 1: Correlations between the posterior median of $\mathbf{m}_j (= \mathbf{f}_j + \mathbf{g}_j)$ and the actual realization for each of the three equations in the DGP.

	\mathbf{m}_1	\mathbf{m}_2	\mathbf{m}_3
Avg.	0.945	0.929	0.754
SD	0.012	0.020	0.045

Notes: Results are obtained from simulating 100 realizations from the proposed DGP. **Avg.** refers to the average correlation, while **SD** to the standard deviation of correlations across realizations.

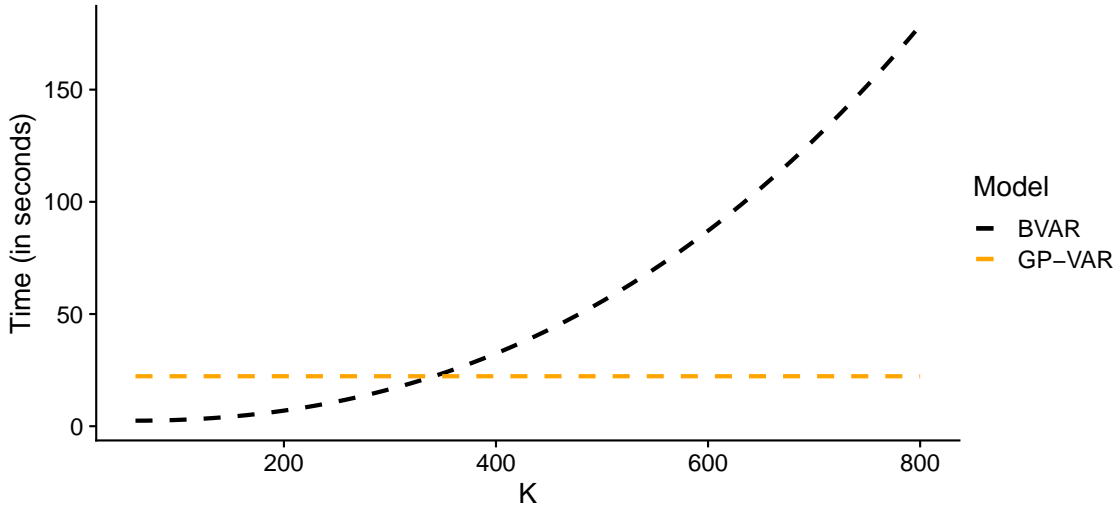
latent components reveals that most of this strong fit is driven by accurate estimates of \mathbf{f}_j . Considering the estimates for \mathbf{g}_j suggests that, if the actual relationship between \mathbf{z}_t and \mathbf{y}_t is non-existent, our approach accurately detects this behavior. However, there are some cases where \mathbf{f}_j soaks up variation in \mathbf{g}_j . This, however, does not impact the mean estimate \mathbf{m}_j .

Since this discussion has been based on a single draw from the DGP, one might ask whether the strong performance of the GP-VAR is due to a particularly favorable realization from the DGP. To briefly investigate whether this is the case we show, in the first row of Table 1, the average correlations between the true and posterior median of the functions for 100 draws from the DGP. Numerical standard errors (across these 100 repetitions) are shown in the second row. The first row indicates that correlations are high, reaching 0.95 for the first, 0.93 for the second and 0.75 for the third equation. The fact that mean correlations slightly decrease are mainly driven by the fact that equations two and three feature substantial high frequency movements which, in our framework, is mostly picked up by the stochastic volatility component.

One key advantage of the GP-VAR is that we can remain agnostic on the form of non-linearities the conditional mean function might take. This is confirmed for synthetic data where, irrespective of the non-linear nature of the model, the estimated sum of \mathbf{f}_j and \mathbf{g}_j (i.e., \mathbf{m}_j) closely tracks the dynamics of the actual outcome. This finding holds both for the linear case (i.e., the first equation), even with fat-tailed errors, and for highly non-linear situations (i.e., the second and third equations).

We have stressed that our approach is computationally efficient and scalable to large datasets. To investigate this claim more carefully, Figure 4 shows the time required to generate 1000 draws from the joint posterior for a given equation across different values of K and for $T = 200$. We show the computation times for our GP-VAR and a VAR with SV. Since our approach is embarrassingly parallel the actual times for generating a

Figure 4: Computation time for 1000 draws from the joint posterior distribution.



Notes: Computation time for 1000 draws from the joint posterior distribution of a standard BVAR (black) and the GP-VAR (orange). Since our approach can be parallelized, the actual times for generating 1000 draws from the joint posterior of the full system is approximately M times the runtimes reported in this figure. Computation times are based on a desktop machine with an AMD Ryzen 7 5800X 8-Core processor.

draw from the joint posterior of the full system are approximately M times the runtimes reported in the figure.¹²

The most striking take away from the figure is that the computation time of the GP-VAR does not depend on K . This implies that increasing the number of lags and/or endogenous variables does not impact estimation times considerably. By contrast, the time necessary to generate a draw from the joint posterior of the VAR rises rapidly in K . This shows that our approach scales well in high dimensions and, in fact, is much faster than competing approaches to non-linear VAR models such as TVP-VARs or regime-switching VARs.

To provide a rough gauge on actual estimation times for practitioners, MCMC estimation of the GP-VAR with eight endogenous variables and five lags takes around 30.5 minutes on a standard desktop computer (for 10,000 MCMC draws). These are gross estimation times and thus include pre-computation of matrices used during MCMC simulation and are not based on parallel computation of the individual equations (which is possible due to the structural form). Estimating larger models (such as the 64 variable GP-VAR) takes around four hours.

¹²Since we need to augment the j^{th} equation with the contemporaneous values of the preceding $j - 1$ equations, this statement is only approximately valid.

5 Modeling the US economy using GP-VARs

In this section, we apply our GP-VAR to US macroeconomic data. Sub-section 5.1 provides details on the data and Sub-section 5.2 discusses whether the GP-VAR fits the data well, shows some of its key in-sample features, and investigates its forecast performance.

5.1 Data overview

We use the quarterly version of the dataset proposed in [McCracken and Ng \(2016\)](#) and consider time series that range from 1960Q1 to 2019Q4. We exclude the years of the Covid-19 pandemic to make the comparison with a linear VAR similar to that used in [Jurado, Ludvigson, and Ng \(2015\)](#) (JLN) sensible. In the following, we consider four different specifications that differ in the number of endogenous variables used. These specifications are:

- GP-VAR-8: This dataset is patterned after the original JLN dataset. It includes the JLN macroeconomic uncertainty index (labeled as UNC), real GDP (RGDP), civilian employment (EMP), average weekly working hours in manufacturing (AWH), consumer price index (CPI), average hourly earnings in manufacturing (AHE), Fed funds rate (FFR), and the S&P 500 (SP500).
- GP-VAR-16: In addition to the variables in the VAR with $M = 8$ endogenous variables, we also include the components of GDP (such as real personal consumption and real private fixed investment), additional labour market variables (such as unemployment, initial claims, average weekly working hours and average hourly earnings across all sectors), housing starts, as well as the real M2 money stock.
- GP-VAR-32: On top of the variables of the VAR with $M = 16$ variables, we include important financial variables, further data on housing and data on loans.
- GP-VAR-64: The largest model we consider features $M = 64$ endogenous variables. The set of endogenous variables is obtained by taking the variables with $M = 32$ and including additional financial variables and data on manufacturing.

All variables are transformed to be approximately stationary and we include five lags of the endogenous variables. The precise variables included (and transformations applied to

each variable) are shown in Table B.1 in the Online Appendix. We consider these different model sizes for several reasons. First, we would like to assess how adding additional information impacts the forecasting performance and the responses of key variables to an uncertainty shock. Second, we are interested in the relationship between non-linearities and the size of the model.

5.2 Predictive evidence and in-sample features

In this section, we start by providing some predictive evidence of our GP-VAR and investigate how the role of the parameters associated with the kernel impact predictive accuracy. To this end, we employ a recursive forecasting design. Our initial training period goes from 1960Q1 to 1999Q4. After computing the one- and four-step-ahead predictive distributions, we add an additional observation, recompute all models and simulate from the corresponding predictive densities. This procedure is repeated until the end of the sample (2019Q4) is reached.

We consider three ways of specifying the equation-specific ($j = 1, \dots, M$) and kernel-specific ($k = \{1, 2\}$) hyperparameters. The first one is a semi-automatic approach that is based on putting the (inverse) length scale and the linear scaling parameter on a two dimensional grid $\kappa_{jk} \in [0.1\bar{\kappa}_{jk}, 2\bar{\kappa}_{jk}]$ and $\xi_{jk} \in [0.04, 4]$. Notice that the median heuristic is used to determine the lower and upper bound of the grid. The second approach (labeled “semi-automatic w/o linear scaling” in Table 2) puts $\kappa_{jk} \in [0.1\bar{\kappa}_{jk}, 2\bar{\kappa}_{jk}]$ on a grid and sets $\xi_{jk} = 1$. Finally, we also consider a specification that does not rely on the median heuristic (labeled “naive” in the table). The grid is $\kappa_{jk} \in [0.1, 2]$ and ξ_{jk} is again set equal to 1.

As competing models, we include the standard Minnesota BVAR with SV, a TVP-VAR-SV similar to the one used in Primiceri (2005) and the BART-VAR with SV proposed in Huber and Rossini (2022). All models are estimated for different model sizes and benchmarked to a small-scale Minnesota BVAR with SV for the three variables we focus on: real GDP growth (RGDP), CPI inflation (CPI), and the Fed funds rate (FFR). For the three focus variables, we compute log predictive Bayes factors (LPBFs) relative to the small-scale BVAR with SV, so that positive numbers indicate that a given model works better than the benchmark while negative values suggest a weaker forecasting performance. The LPBFs do not only take into account how well a given model predicts the

Table 2: Log predictive Bayes factors (LPBFs) relative to a small-scale BVAR with stochastic volatility (SV).

Specification		One-step-ahead				Four-step-ahead			
M	Grid for kernel hyperparameters	Joint	RGDP	CPI	FFR	Joint	RGDP	CPI	FFR
GP-VAR homoskedastic									
64	semi-automatic	-90.310	-7.769	-12.312	-59.702	-52.285	1.021	-9.443	-31.713
	semi-automatic w/o linear scaling	-78.323	-1.832	-14.770	-49.583	-45.304	5.940	-10.795	-22.169
	naive	-84.216	-4.953	-23.869	-53.582	-38.794	3.785	-4.312	-24.104
32	semi-automatic	-82.036	-8.549	-11.667	-58.767	-45.526	1.816	-6.588	-30.567
	semi-automatic w/o linear scaling	-72.908	-2.449	-16.077	-48.757	-53.405	4.485	-22.694	-17.920
	naive	-71.102	-3.640	-16.168	-51.738	-46.112	2.980	-14.723	-20.679
16	semi-automatic	-80.595	-7.388	-11.332	-58.063	-45.852	1.234	-6.506	-30.582
	semi-automatic w/o linear scaling	-83.747	-0.300	-18.696	-45.319	-51.862	2.849	-20.447	-18.794
	naive	-81.852	0.088	-15.936	-47.053	-42.958	2.051	-10.318	-20.849
8	semi-automatic	-75.047	-7.007	-11.688	-52.163	-43.450	3.441	-6.034	-27.388
	semi-automatic w/o linear scaling	-78.854	-1.254	-21.087	-46.306	-46.022	4.743	-15.341	-19.580
	naive	-73.726	-2.155	-16.128	-46.877	-43.949	4.520	-12.480	-22.536
GP-VAR SV									
64	semi-automatic	-7.995	-5.423	-0.426	-2.106	6.370	-6.907	11.058	9.543
	semi-automatic w/o linear scaling	-15.400	-6.154	-4.971	-1.117	-0.077	-14.521	7.616	8.992
	naive	-4.490	-5.829	3.954	0.186	-4.247	-9.763	9.384	8.585
32	semi-automatic	-3.673	-4.881	-0.672	3.699	10.102	-4.945	12.764	7.434
	semi-automatic w/o linear scaling	-5.219	-5.661	-0.629	2.134	-0.016	-12.096	11.029	4.948
	naive	-5.751	-5.225	-0.902	3.905	3.513	-12.650	11.346	5.793
16	semi-automatic	10.206	-4.387	7.387	7.222	17.899	-7.682	15.316	10.967
	semi-automatic w/o linear scaling	4.986	-4.640	6.485	6.010	7.336	-14.249	18.058	9.332
	naive	7.444	-2.179	6.436	9.980	6.735	-13.909	15.195	8.764
8	semi-automatic	4.459	-7.142	10.887	3.285	7.989	-9.383	16.009	8.377
	semi-automatic w/o linear scaling	7.496	-7.652	12.122	7.648	13.661	-10.402	14.084	9.753
	naive	8.994	-7.158	11.004	11.154	8.808	-11.680	14.594	9.055
BART-VAR SV									
64		-41.432	-5.128	-12.146	-18.671	-10.984	2.714	1.216	-11.179
32		-39.967	-5.513	-4.017	-26.650	-17.564	-3.351	5.918	-15.872
16		-32.691	-2.987	-2.448	-22.042	-17.153	0.580	1.398	-17.512
8		-33.579	-5.345	-3.765	-20.196	-13.522	5.349	0.769	-21.598
Minnesota TVP-VAR SV									
8		1.648	-1.531	2.220	-2.268	-3.543	-1.877	0.948	-0.359
Minnesota BVAR SV									
64		-0.220	-0.347	-8.011	8.268	-23.563	-3.944	-18.503	-2.696
32		-7.983	2.592	-9.862	0.911	-12.071	-4.846	-6.127	2.137
16		3.426	2.296	1.746	-0.003	-5.153	-1.373	-2.152	0.137
8	Benchmark	-223.987	-73.554	-138.633	-11.961	-278.520	-91.359	-151.355	-41.429

Notes: The table shows joint and marginal LPBFs relative to a small-scale BVAR with SV for the one- and four-step-ahead horizons. The first column indicates model size, the second column provides information on how the hyperparameters are treated. We consider three cases: “semi-automatic” refers to a two-dimensional grid for both the (inverse) length scale and the linear scaling parameter which is scaled using the median heuristic, “semi-automatic w/o linear scaling” to a grid only for the (inverse) length scale and sets $\xi_{jk} = 1$ while “naive” refers to a grid only for the (inverse) length scale without scaling the grid using the median heuristic and $\xi_{jk} = 1$ ($j = 1, \dots, M$; $k = \{1, 2\}$). Positive values imply that a particular model improves upon the small-scale BVAR with SV, while negative values suggest that the benchmark is preferred. The best specification for each variable and horizon combination is shown in bold. The red shaded entries associated with the benchmark give the actual log predictive likelihoods (LPLs).

realization of a given variable but also factor in forecasting performance for higher order features of the predictive distribution (for a discussion, see Geweke and Amisano, 2010). To investigate whether controlling for heteroskedasticity pays off in the GP-VAR, we also consider homoskedastic variants of the GP-VAR in the upper part of the table.

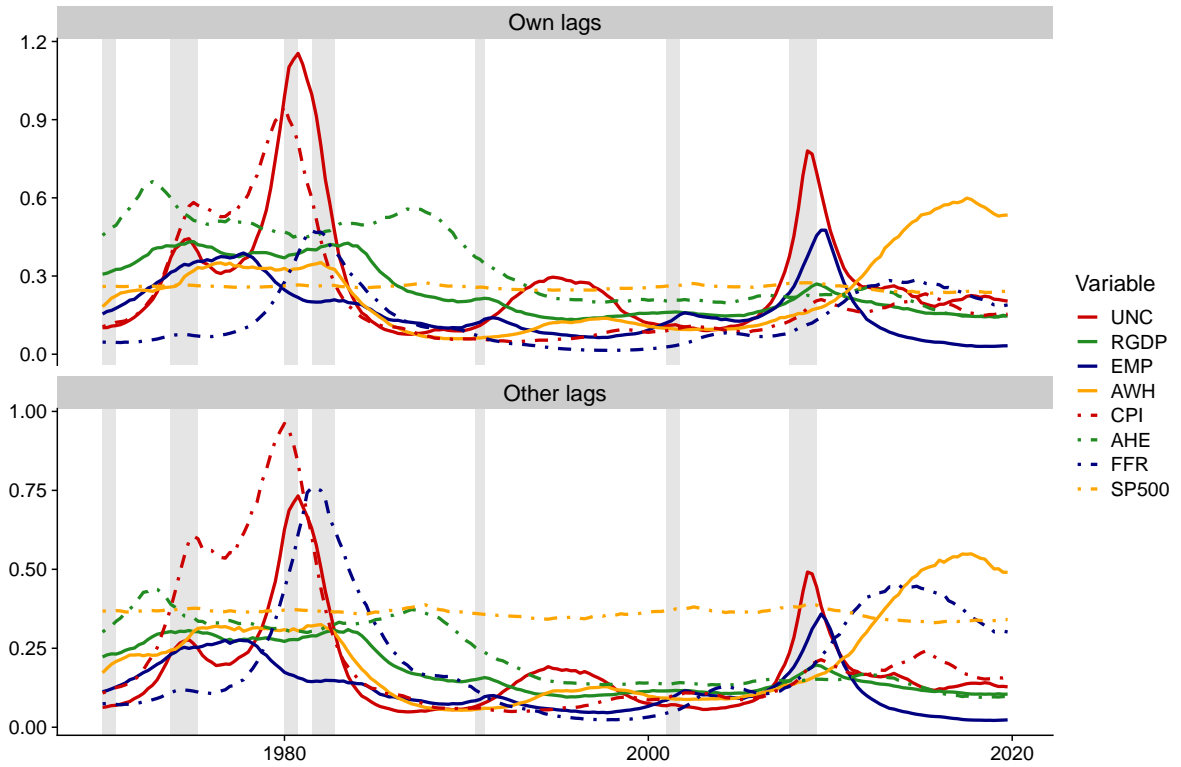
Table 2 shows the results of our forecasting exercise across different model sizes. The columns “Joint” show the joint LPBF for the three focus variables and thus provide

a comparable (across model sizes) metric of overall predictive accuracy. Considering joint LPBFs reveals that GP-VAR SV with $M = 16$ and a semi-automatic approach for hyperparameter elicitation improves upon all competing models for both forecast horizons. The smallest model ($M = 8$) also yields competitive predictions. Once we further increase the size of the dataset, predictive accuracy slightly deteriorates. Interestingly, and consistent with findings in Clark, et al. (2021), we find that the gains in predictive accuracy increase when we focus on higher forecast horizons. Comparing the models with SV to their homoskedastic counterparts paints a very consistent picture. Models which do not control for time variation in the error variances perform consistently worse than the models that have SV in the error terms.

To drill deeper into which variables drive the overall forecasting performance, we now focus on the marginal LPBFs for the three focus variables. Starting with one-quarter-ahead predictions of GDP growth, we observe that the GP-VARs with SV are beaten by the BVAR-SV with $M = 32$. When we turn off SV, predictive accuracy sometimes increases by small margins. This result, however, changes if we focus on higher order forecasts. For one-year-ahead predictions of GDP growth, the single best performing model is the largest ($M = 64$) homoskedastic GP-VAR with the semi-automatic approach that fixes the linear scaling parameter to one. Strikingly, at that horizon and for this specific variable, using a homoskedastic specification is almost uniformly better than the corresponding SV setup. For inflation and the Fed funds rate, the smaller-sized GP-VARs again yield the best density forecasting performance, outperforming all competing models. Finally, focusing on interest rate forecasts shows that GP-VARs with SV do well (for all model sizes) but once we turn off SV forecasts become highly imprecise. This is driven by the fact that during the zero lower bound, the conditional variance of the interest rate equation approaches zero and a model which assumes homoskedasticity fails to take that into account.

After having established that the different GP-VARs with SV do well when used to forecast US macroeconomic quantities, we focus on some in-sample features for the GP-VAR with eight endogenous variables. To get an impression on how the linear shrinkage parameter evolves over time, Figure 5 plots the product of the error variances times the linear shrinkage parameter that determines the kernel of \mathbf{f}_j and \mathbf{g}_j . Two interesting features emerge. First, less shrinkage (larger parameter values) is applied to the own lags

Figure 5: Linear shrinkage parameters of equation-specific kernels for the GP-VAR-8.

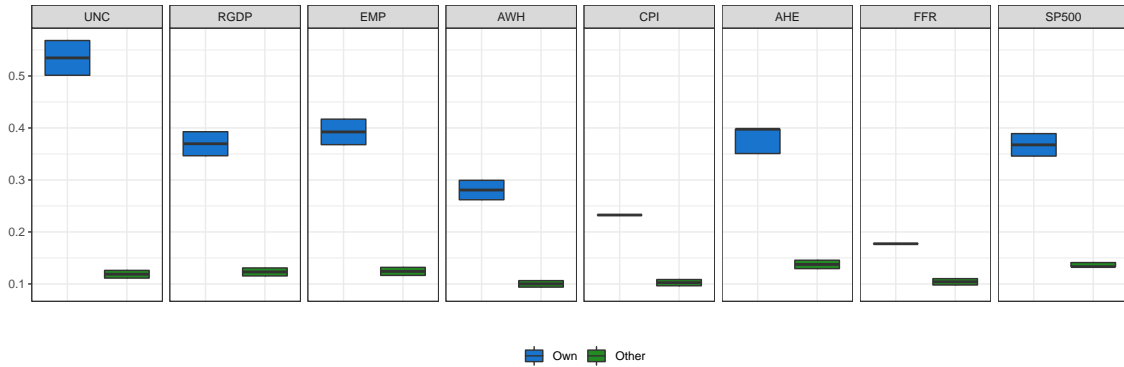


Notes: This figure reports the posterior means of the product of the error variances ω_{jt} and the linear scaling parameters for own lags (ξ_{j1}) and for other lags (ξ_{j2}), respectively. These two quantities correspond to the diagonal elements of the re-scaled kernels $\omega_{jt} \times k_{\vartheta_{j1}}(\mathbf{x}_t, \mathbf{x}_t) = \omega_{jt}\xi_{j1}$ and $\omega_{jt} \times k_{\vartheta_{j2}}(\mathbf{z}_t, \mathbf{z}_t) = \omega_{jt}\xi_{j2}$.

than to the lags of the other variables. This holds for most variables under scrutiny (except for CPI inflation and S&P 500 returns). Second, less shrinkage is typically applied during recessionary times. In particular, we introduce little shrinkage during the recessions in the early 1980s and the financial crisis. Notice, however, that there are also some exceptions from this pattern (such as stock market returns, hours worked or hours employed). This finding is, again, in line with results from the forecasting literature showing that more information is particularly useful during problematic times (see, e.g., Koop, 2013).

Finally, we investigate differences in κ_{j1} and κ_{j2} across equations ($j = 1, \dots, M$) and variable types. Boxplots that show the posterior distribution of the inverse of the length scale parameters for own and other lags are in Figure 6. Recall that large values of κ_{jk} ($k = 1, 2$) imply more variation in the latent processes whereas values of κ_{jk} close to zero imply less variation in \mathbf{f}_j and \mathbf{g}_j . A general pattern is that for all variables the hyperparameters associated with the kernel on own lags are considerably larger than the ones for the kernel related to the other lags. This indicates that the own lags of a given endogenous variable require more flexibility (i.e., functions that allow for much more

Figure 6: Inverse length scale parameters of equation-specific kernels for the GP-VAR-8.



Notes: This figure reports the posterior summaries in the form of simplified boxplots of the inverse length scale parameters for own lags (κ_{j1}) and other lags (κ_{j2}), respectively. The solid black lines denote the posterior medians, while the blue (green) shaded areas represent the 50% posterior credible sets (i.e., the posterior interquartile ranges).

variation) whereas the effect of other lags appears to be more linear. For the majority of variables (except for hours worked, CPI inflation and the Fed funds rate), the posterior distribution of the hyperparameter looks similar. For the three exceptions, κ_{j1} is much smaller and more precisely estimated.

6 The macroeconomic effects of uncertainty shocks

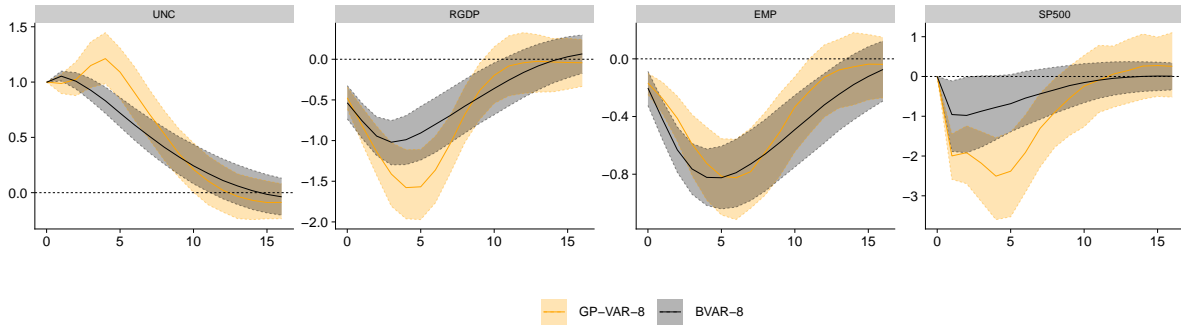
We now analyze the effects of uncertainty shocks using our GP-VAR and focus on assessing how macroeconomic uncertainty feeds through the economy. In Sub-section 6.1, as a benchmark exercise, we compare our impulse responses to those obtained from a model similar to that used by JLN. In Sub-section 6.2 we leverage the non-linear nature of the GP-VAR and analyze how the effects of uncertainty shocks change according to the sign or size of the shocks, and over time.

6.1 Comparison with standard BVAR analysis

We benchmark the IRFs of our GP-VAR-8 to the ones of a BVAR with SV that is closely related to the original JLN specification.¹³ In what follows, our focus will be on the variables discussed in JLN: year-on-year growth rates of output (measured through real GDP) and employment. We also show the responses of the uncertainty indicator and the quarter-on-quarter returns of the S&P 500 and include the responses of the other variables in Section C of the Online Appendix.

¹³While they use a classical homoskedastic VAR estimated on monthly data, we work with a quarterly BVAR with SV.

Figure 7: Impulse responses of focus variables in the GP-VAR-8 relative to a small-scale BVAR.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. GP-VAR-8 refers to the smallest variant of our non-parametric model and BVAR-8 refers to a small-scale BVAR with SV, which is closely related to the specification used in [Jurado, Ludvigson, and Ng \(2015\)](#).

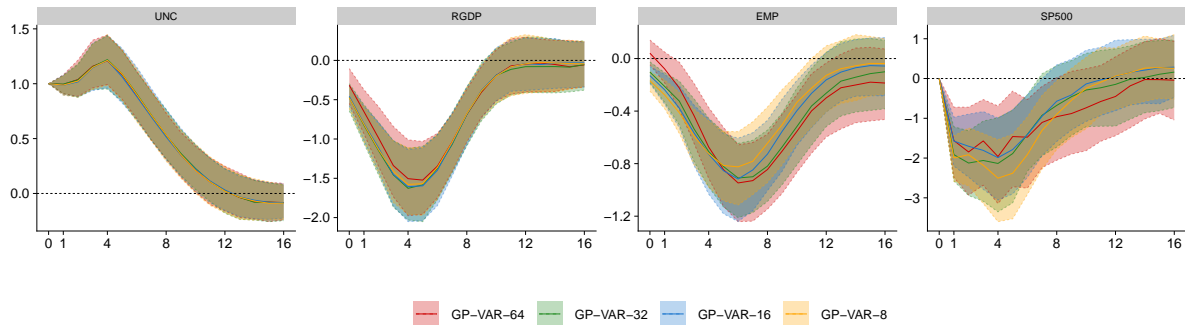
In [Figure 7](#) we report the (average over time in the case of the GP-VAR) posterior quantiles (16th, 50th and 84th) of the responses to a macroeconomic uncertainty shock in the JLN model (in gray) and in the corresponding GP-VAR (in orange) with eight endogenous variables and uncertainty ordered second after stock market returns. Uncertainty responses to its own shock differ slightly between the GP-VAR and the linear model. These differences relate to responses within the first five quarters after the shock hit the system. The BVAR yields uncertainty reactions that peak after one quarter, declining steadily afterwards. As opposed to this swift reaction in uncertainty, the GP-VAR generates endogenous uncertainty reactions which peak after five quarters, declining sharply afterwards. After around eight quarters, both IRFs (almost) coincide.

This uncertainty reaction has direct implications on how the other variables in the model react. Real GDP growth reacts in an hump-shaped manner under both models. However, driven by the somewhat later peak in uncertainty, the GP-VAR produces much stronger output growth reactions that peak slightly later (after around five quarters). When we focus on employment growth the IRFs differ less. In principle, both models suggest a peak decline of around 0.8 percentage points, with the GP-VAR generating a somewhat slower response, reaching its trough after about six to seven quarters. But in principle, responses between the linear and non-parametric model tell a similar story. Finally, financial market reactions measured through the S&P500 suggest a much stronger decline in stock prices under the GP-VAR. Interestingly, the shape of the IRFs suggests that the linear model generates the strongest reaction after around two quarters. In the GP-VAR, we find that stock markets react faster and stronger to uncertainty shocks, with

substantial reactions within the first year after the shock hit the system.

To conclude, in Figure 8 we report the GIRFs to the uncertainty shock for the same four variables displayed in Figure 7 but obtained from GP-VARs of different dimensions (with 8, 16, 32, and 64 variables). Differences across model sizes are small (or non-existent) for most variables. Small differences arise for employment growth, with the magnitude of the responses increasing with the model size. Stock market reactions also differ slightly across datasets, with no clear-cut pattern. Since the GIRFs are very similar across model size and given its excellent forecasting properties, we will focus on asymmetries generated by the GP-VAR-8 model in the following sections. Results for the larger models are provided in the Online Appendix.

Figure 8: Impulse responses of focus variables across different information sets.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive one standard deviation shock in macroeconomic uncertainty across different information sets. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets.

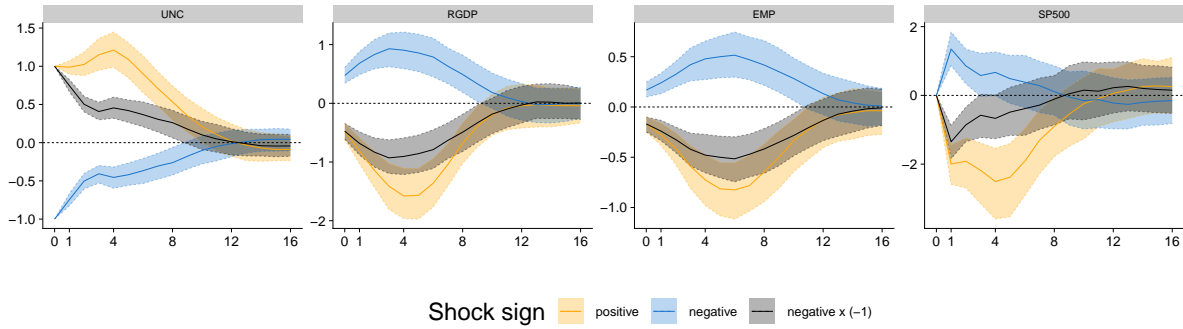
6.2 Asymmetries in the transmission of uncertainty shocks

The non-linear and non-parametric nature of our models allows for asymmetries in the impulse response functions. This implies that shocks propagate non-linearly through the model, giving rise to differences in the GIRFs both over time but also for different shock magnitudes or signs.

6.2.1 Asymmetries with respect to the sign of the shock

The first aspect we consider relates to whether positive and negative uncertainty shocks trigger different responses of the economy. In Figure 9 we report the responses to negative and positive uncertainty shocks from the GP-VAR-8, averaged over time. The figure thus shows GIRFs to a positive (in orange), negative (in blue) and a negative shock multiplied

Figure 9: Shock sign asymmetries in responses of focus variables for the GP-VAR-8.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a negative (positive) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, negative $\times(-1)$ denotes a negative one standard standard deviation shock with the respective responses being mirrored across the x-axis.

by -1 (in gray, to ease comparison). From the figure we observe some differences. These differences mostly relate to peak reactions as well as short-run (i.e. within two years) responses. In general, we find that positive shocks (higher uncertainty) trigger a stronger reaction of uncertainty, which in turn translates into more pronounced reactions of real activity and stock market quantities.

More specifically, considering the endogenous reaction of the uncertainty indicator shows that responses to a positive uncertainty shock peak after around a year and quickly die out afterwards. However, if uncertainty unexpectedly declines, the peak happens on impact and is much smaller as opposed to an adverse uncertainty shock.

Turning to real GDP and employment growth, we find that positive shocks trigger stronger reactions for both variables. Interestingly, the timing of the peak responses is similar for negative and positive shocks but reactions appear much more pronounced for the latter. Stock market reactions also differ markedly across positive and negative shocks. For positive shocks we, again, find that the peak effect happens after one year and that it is more pronounced as compared to the negative shock. Overall, the picture that emerges from the GP-VAR is that higher unexpected uncertainty has stronger effects on the economy than lower uncertainty, a feature that is a priori ruled out in linear VARs.

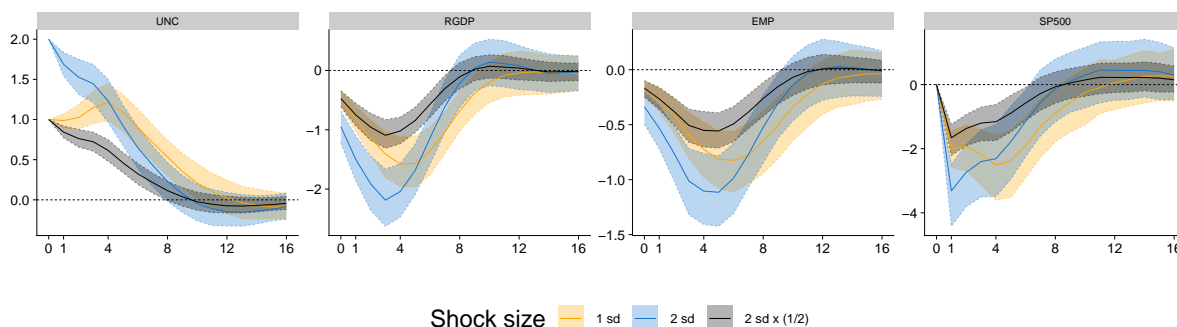
6.2.2 Asymmetries with respect to the size of the shock

Our GP-VAR also allows for analyzing how shocks of different sizes impact the economy. As opposed to a standard VAR which assumes that shocks enter linearly (and thus responses to shocks of different sizes are exactly proportional to each other) our GP-VAR is more flexible and allows for investigating whether shocks of different magnitudes trigger

different dynamics in the GIRFs.

In Figure 10 we consider two shock sizes: a one standard deviation and a two standard deviation shock. To permit straightforward comparison of the shapes of the responses to differently sized shocks, we also add impulses to a two standard deviation shock which are then re-scaled to match the impact of the one standard deviation shock (the gray shaded area in the figures).

Figure 10: Shock size asymmetries in responses of focus variables for the GP-VAR-8.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive two (one) standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, 1 sd refers to a one standard deviation shock, 2 sd indicates a two standard deviation shock, and $2 \text{ sd} \times (1/2)$ denotes a two standard deviation shock with the respective responses divided by two.

This figure gives rise to at least two observations. First, when we compare the shape of the responses to a one standard deviation to the ones of a two standard deviation shock we find differences in the timing (and more generally in the shape) of the IRFs. A stronger shock triggers a faster peak reaction of GDP and employment growth. Stock market reactions display a somewhat different shape. After a sharp immediate reaction (for both shock sizes) the peak effect happens to be on impact if the size of the shock is large whereas it turns out to materialize after one year if the shock size is smaller.

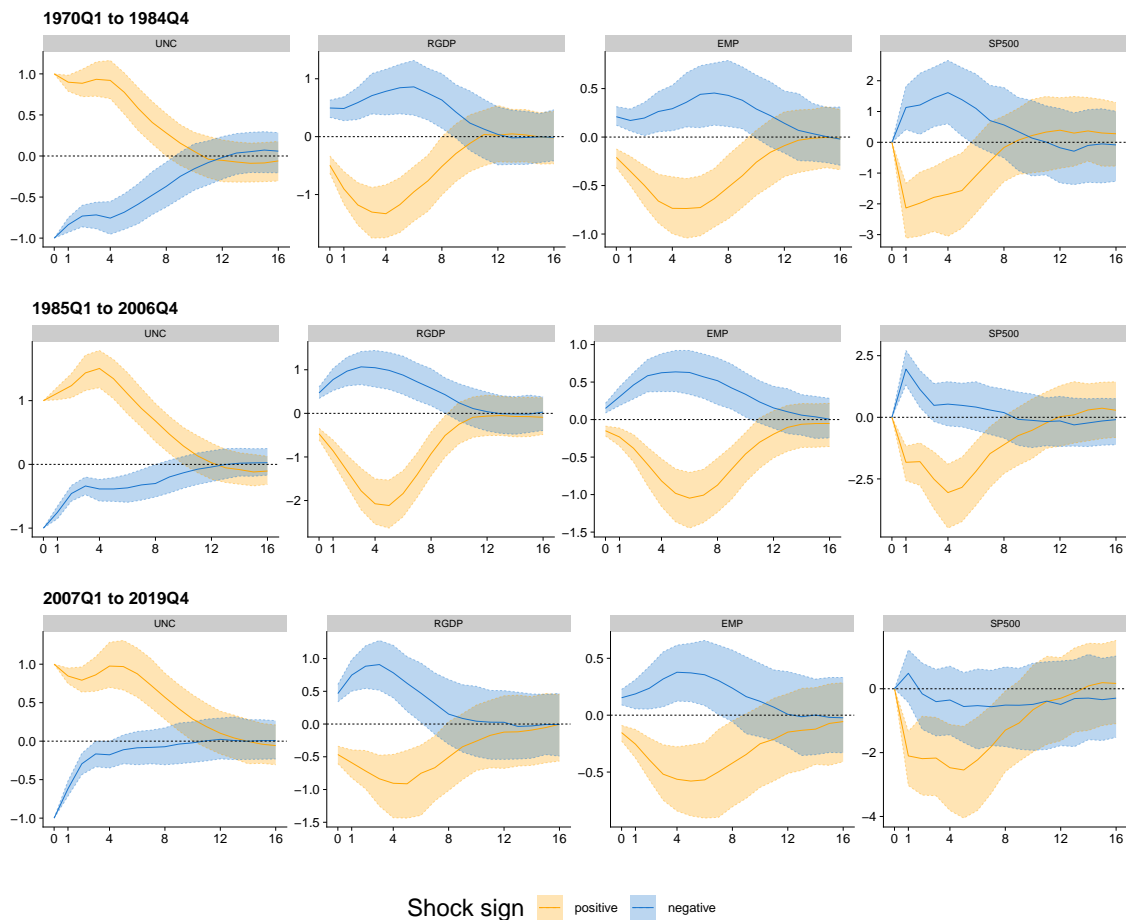
Second, in terms of the magnitudes we find that a two standard deviation shock triggers peak responses with magnitudes that are less than twice the magnitudes to a one standard deviation shock. This is particularly visible for employment and output reactions. For stock market responses, the impact reactions are (almost) proportional to each other.

6.2.3 Asymmetries over time

After showing that the economy reacts asymmetrically with respect to the sign and size of the uncertainty shock, this section asks whether the effect of uncertainty shocks changes

over time (see Section 2.5 of [Castelnuovo \(2022\)](#), or [Mumtaz and Theodoridis \(2018\)](#) for some evidence using TVP-VARs).

Figure 11: Impulse responses of focus variables in the GP-VAR-8 across different sub-sample periods.



Notes: Period-specific average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive (negative) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets.

We start by considering impulse responses averaged over certain sub-periods in [Figure 11](#). The classification into sub-periods is mostly taken from [D’Agostino and Surico \(2012\)](#), and it is such that the main events in each sub-period include, respectively, the great inflation (1970Q1 to 1984Q4), the great moderation (1985Q1 to 2006Q4), and the post great moderation period (2007Q1 to 2019Q4). To also get a rough feeling about whether asymmetries between positive and negative shocks have changed over time, all figures include the IRFs to positive (in orange) and negative (in blue) shocks.

The main feature emerging from the figure is the different behavior of the response of uncertainty across sub-samples. In the final two sub-samples, uncertainty responses increase up to four quarters after the shock, with peak effects being strongest in the great

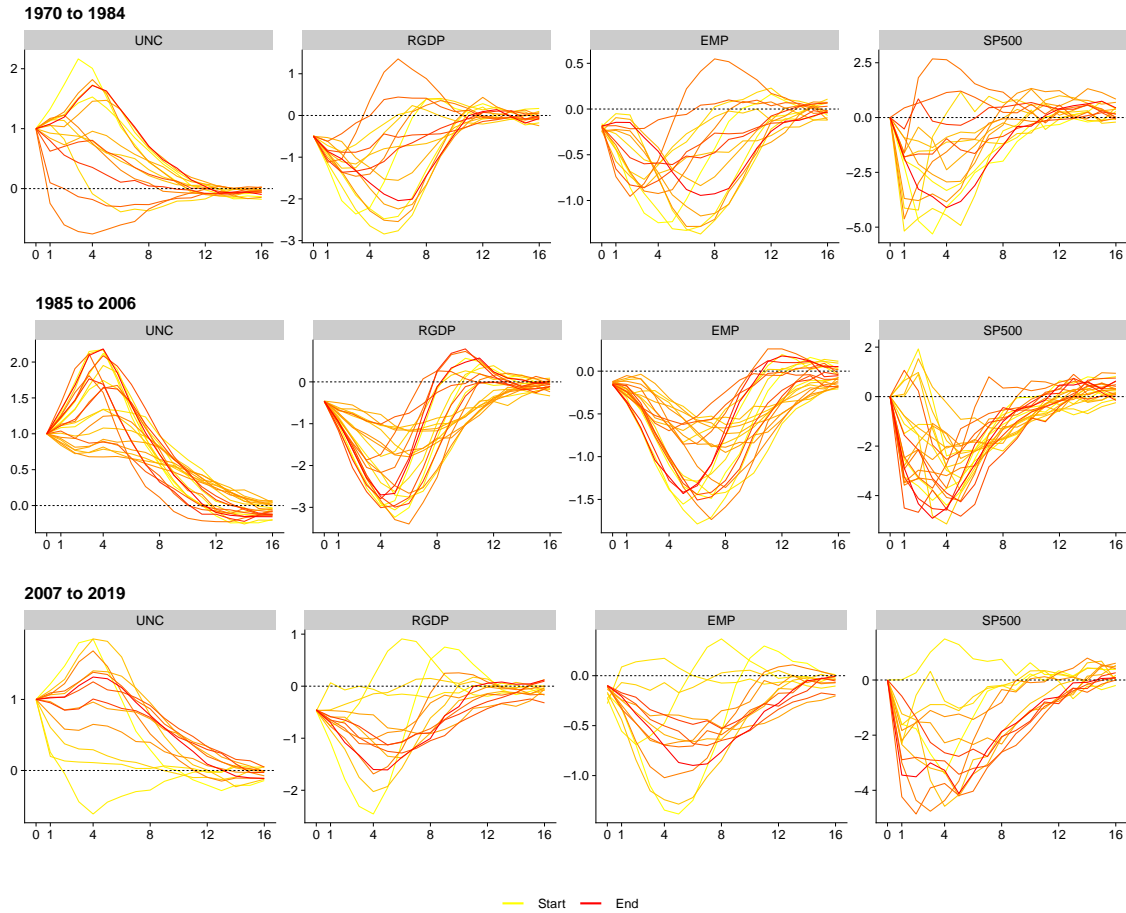
moderation period and becoming slightly weaker in the final sub-sample. Moreover, sign effects of uncertainty responses increase appreciably in the last two sub-samples.

These differences in the responses of uncertainty trigger differences in the IRFs of the other quantities which relate not only to the magnitudes but also to the shapes of the responses. We find that GDP growth, employment and the S&P 500 display the strongest reactions in the great moderation regime, becoming slightly weaker in the post great moderation period. The weaker reaction of real activity over time corroborates findings in [Mumtaz and Theodoridis \(2018\)](#) who also report smaller responses of real activity to uncertainty shocks. As opposed to their findings, we observe that stock market reactions do not change much in magnitude but the shape differs (in accordance with the different shape in the uncertainty reaction described above). We, moreover, observe that asymmetries in terms of the sign of the shocks have decreased over time for GDP and employment growth. Only for stock market reactions these sign asymmetries have increased, with benign uncertainty shocks yielding a much weaker positive reaction of stock markets during the post great moderation regime.

Finally, to conclude this section we raise the issue that considering GIRFs averaged over sub-samples possibly still masks important differences over time within sub-samples. To shed light on whether IRFs change within regimes, [Figure 12](#) displays yearly averages of posterior medians of the IRFs over time during each of the three periods. Yellow IRFs refer to the beginning of the respective sub-sample and red ones denote IRFs computed towards the end of the sub-sample. This figure suggests substantial heterogeneity in responses during the great inflation period. Especially towards the end of this sample, reactions of GDP growth and employment point towards a substantial real activity overshoot. During the great moderation, the intra-period variation of the IRFs becomes much smaller, yielding patterns more consistent with the common wisdom in the uncertainty literature: real activity and stock markets decline in response to increases in economic uncertainty. In the years from 2007 to 2019, we find that IRFs differ especially in the beginning of the sample (from 2007 to 2009). For the remaining years, there is much less variation in responses and these appear to be similar to those observed in the 1985 to 2006 period.

Overall, we can conclude that the effects of uncertainty change both during sub-samples defined by economic considerations and sometimes also within each sub-sample.

Figure 12: Period-specific impulse responses of focus variables in the GP-VAR-8 across different sub-sample periods.



Notes: Impulse response functions to a positive one standard deviation shock in macroeconomic uncertainty in the GP-VAR-8 across sub-sample periods. Solid lines denote the yearly averaged posterior medians, with colors ranging from yellow (start of the sample) to red (end of the sample).

This kind of time variation is a priori ruled out in linear VAR models, which can therefore lead to biased estimates of the effects of uncertainty.

7 Conclusions

In this paper, we have developed a flexible multivariate model that uses Gaussian processes to model the unknown relationship between a panel of macroeconomic time series and their lagged values. Our GP-VAR is a very flexible model which remains agnostic on the precise relations between the endogenous variables and the predictors. This model can be viewed as a very flexible and general extension of the linear VAR commonly used in empirical macroeconomics. We also control for changes in the error variances by introducing a stochastic volatility specification. While a more flexible conditional mean can reduce the

need of a time-varying conditional variance, empirically we find heteroskedasticity to be relevant also for GP-VARs.

We develop efficient MCMC estimation algorithms for the GP-VAR, which are scalable to high dimensions, so much so that for large models estimation is even faster than for the corresponding BVAR-SV. Scaling the covariance of the Gaussian process by the latent volatility factors is particularly helpful to achieve computational gains, as it permits to pre-compute several quantities before MCMC sampling. This speeds up computation enormously.

To illustrate the practical working of the GP-VAR, we first test it on simulated data from different linear and non-linear models, finding that it is capable of reproducing a variety of non-linear patterns (but also a linear behavior). Then, we show in a forecasting exercise that our model yields favorable density forecasts of US output, inflation and short-term interest rates with respect to both linear and other non-parametric and time-varying specifications.

In the main part of our empirical work we re-assess the effects of uncertainty shocks by replicating and extending the analysis carried out by [Jurado, Ludvigson, and Ng \(2015\)](#) based on linear VARs with the GP-VAR. Overall, our empirical results suggest that the measurement of uncertainty and its effects with a simple linear VAR can lead to several incorrect conclusions. Not only the effects of uncertainty can be over-stated, but they can also be treated as stable over time, symmetric for positive and negative shocks, and proportional to the shock size. Instead the GP-VAR model, which is preferred to the linear VAR in terms of fit and forecasting performance, returns time variation in the responses, asymmetry and non-proportionality. Hence, the empirical features we uncover should be also replicated by theoretical models about uncertainty and its effects, which instead at the moment typically assume stability and symmetry (see, e.g., the survey in [Bloom, 2014](#)).

References

- Aastveit, Knut Are, Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino (2017), “Have standard VARs remained stable since the crisis?” *Journal of Applied Econometrics*, 32, 931–951.
- Aastveit, Knut Are, Gisle James Natvik, and Sergio Sola (2017), “Economic uncertainty and the influence of monetary policy,” *Journal of International Money and Finance*, 76, 50–67.
- Alessandri, Piergiorgio, and Haroon Mumtaz (2019), “Financial regimes and uncertainty shocks,” *Journal of Monetary Economics*, 101, 31–46.
- Bañbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010), “Large Bayesian vector auto regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- Bassetti, Federico, Roberto Casarin, and Fabrizio Leisen (2014), “Beta-product dependent Pitman–Yor processes for Bayesian inference,” *Journal of Econometrics*, 180, 49–72.
- Billio, Monica, Roberto Casarin, and Luca Rossini (2019), “Bayesian nonparametric sparse VAR models,” *Journal of Econometrics*, 212, 97–115.
- Bloom, Nicholas (2014), “Fluctuations in uncertainty,” *Journal of Economic Perspectives*, 28, 153–76.
- Caggiano, Giovanni, Efrem Castelnuovo, and Giovanni Pellegrino (2017), “Estimating the real effects of uncertainty shocks at the zero lower bound,” *European Economic Review*, 100, 257–272.
- (2021), “Uncertainty shocks and the great recession: Nonlinearities matter,” *Economics Letters*, 198, 109669.
- Cai, Zongwu, Jianqing Fan, and Qiwei Yao (2000), “Functional-coefficient regression models for nonlinear time series,” *Journal of the American Statistical Association*, 95, 941–956.
- Carriero, Andrea, Todd E Clark, and Massimiliano Giuseppe Marcellino (2022), “Specification choices in quantile regression for empirical macroeconomics,” *FRB of Cleveland Working Paper*, No. 22-25.
- Carvalho, Carlos M, Nicholas G Polson, and James G Scott (2010), “The horseshoe estimator for sparse signals,” *Biometrika*, 97, 465–480.
- Castelnuovo, Efrem (2022), “Uncertainty before and during COVID19: A survey,” *Journal of Economic Surveys*, forthcoming.
- Chan, Joshua C.C. (2017), “The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling,” *Journal of Business & Economic Statistics*, 35, 17–28.
- (2021), “Minnesota-type adaptive hierarchical priors for large Bayesian VARs,” *International Journal of Forecasting*, 37, 1212–1226.
- Chan, Joshua C.C., Gary Koop, Dale J. Poirier, and Justin L. Tobias (2019), *Bayesian Econometric Methods*, Econometric Exercises: Cambridge University Press, 2nd edition.
- Chaudhuri, Arin, Deovrat Kakde, Carol Sadek, Laura Gonzalez, and Seunghyun Kong (2017), “The mean and median criteria for kernel bandwidth selection for support vector data description,” in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 842–849, IEEE.
- Chen, Bin, and Yongmiao Hong (2012), “Testing for smooth structural changes in time series models via nonparametric regression,” *Econometrica*, 80, 1157–1183.
- Chipman, Hugh A, Edward I George, and Robert E McCulloch (2010), “BART: Bayesian additive regression trees,” *The Annals of Applied Statistics*, 4, 266–298.
- Clark, Todd E, Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer (2021), “Tail forecasting with multivariate Bayesian additive regression trees,” *FRB of Cleveland Working Paper*, No. 21-08.
- Cong, Yulai, Bo Chen, and Mingyuan Zhou (2017), “Fast simulation of hyperplane-truncated multivariate Normal distributions,” *Bayesian Analysis*, 12, 1017–1037.
- Coulombe, Philippe Goulet (2020), “The macroeconomy as a random forest,” *arXiv*, 2006.12724.
- (2022), “A neural Phillips curve and a deep output gap,” *arXiv*, 2202.04146.
- Crawford, Lorin, Seth R. Flaxman, Daniel E. Runcie, and Mike West (2019), “Variable prioritization in nonlinear black box methods: A genetic association case study,” *The Annals of Applied Statistics*, 13, 958–989.
- D’Agostino, Antonello, and Paolo Surico (2012), “A century of inflation forecasts,” *Review of Economics and Statistics*, 94, 1097–1106.
- Escobar, Michael D, and Mike West (1995), “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, 90, 577–588.

- Friedman, Jerome H (2001), “Greedy function approximation: a gradient boosting machine,” *Annals of Statistics*, 1189–1232.
- Frühwirth-Schnatter, Sylvia, and Gertraud Malsiner-Walli (2019), “From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering,” *Advances in Data Analysis and Classification*, 13, 33–64.
- Geweke, John, and Gianni Amisano (2010), “Comparing and evaluating Bayesian predictive distributions of asset returns,” *International Journal of Forecasting*, 26, 216–230.
- Giraitis, Liudas, George Kapetanios, and Tony Yates (2014), “Inference on stochastic time-varying coefficient models,” *Journal of Econometrics*, 179, 46–65.
- (2018), “Inference on multivariate heteroscedastic time varying random coefficient models,” *Journal of Time Series Analysis*, 39, 129–149.
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2021), “Autoencoder asset pricing models,” *Journal of Econometrics*, 222, 429–450.
- Hamilton, James D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica*, 357–384.
- Hirano, Keisuke (2002), “Semiparametric Bayesian inference in autoregressive panel data models,” *Econometrica*, 70, 781–799.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989), “Multilayer feedforward networks are universal approximators,” *Neural networks*, 2, 359–366.
- Huber, Florian, and Martin Feldkircher (2019), “Adaptive shrinkage in Bayesian vector autoregressive models,” *Journal of Business & Economic Statistics*, 37, 27–39.
- Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2020), “Nowcasting in a pandemic using non-parametric mixed frequency VARs,” *Journal of Econometrics*, forthcoming.
- Huber, Florian, and Luca Rossini (2022), “Inference in Bayesian additive vector autoregressive tree models,” *The Annals of Applied Statistics*, 16, 104–123.
- Jin, Xin, John M Maheu, and Qiao Yang (2022), “Infinite Markov pooling of predictive distributions,” *Journal of Econometrics*, 228, 302–321.
- Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (2015), “Measuring uncertainty,” *American Economic Review*, 105, 1177–1216.
- Jylänki, Pasi, Jarno Vanhatalo, and Aki Vehtari (2011), “Robust Gaussian process regression with a Student-t likelihood,” *Journal of Machine Learning Research*, 12.
- Kalli, Maria, and Jim E Griffin (2018), “Bayesian nonparametric vector autoregressive models,” *Journal of Econometrics*, 203, 267–282.
- Kapetanios, George, Massimiliano Marcellino, and Fabrizio Venditti (2019), “Large time-varying parameter VARs: A nonparametric approach,” *Journal of Applied Econometrics*, 34, 1027–1049.
- Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), “Stochastic volatility: likelihood inference and comparison with ARCH models,” *The Review of Economic Studies*, 65, 361–393.
- Koop, Gary (2003), *Bayesian Econometrics*: Wiley, Chichester.
- (2013), “Forecasting with medium and large Bayesian VARs,” *Journal of Applied Econometrics*, 28, 177–203.
- Koop, Gary, and Dimitris Korobilis (2013), “Large time-varying parameter VARs,” *Journal of Econometrics*, 177, 185–198.
- Koop, Gary, M. Hashem Pesaran, and Simon M. Potter (1996), “Impulse response analysis in nonlinear multivariate models,” *Journal of Econometrics*, 74, 119–147.
- Kowal, Daniel R, David S Matteson, and David Ruppert (2017), “A Bayesian multivariate functional dynamic linear model,” *Journal of the American Statistical Association*, 112, 733–744.
- Makalic, Enes, and Daniel F Schmidt (2015), “A simple sampler for the horseshoe estimator,” *IEEE Signal Processing Letters*, 23, 179–182.
- McCracken, Michael W., and Serena Ng (2016), “FRED-MD: A monthly database for macroeconomic research,” *Journal of Business & Economic Statistics*, 34, 574–589.
- Mercer, James (1909), “Functions of positive and negative type, and their connection the theory of integral equations,” *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 209, 415–446.
- Mumtaz, Haroon, and Konstantinos Theodoridis (2018), “The changing transmission of uncertainty shocks in the US,” *Journal of Business & Economic Statistics*, 36, 239–252.
- Neal, Radford M (2000), “Markov chain sampling methods for Dirichlet process mixture models,” *Journal of Computational and Graphical Statistics*, 9, 249–265.

- Novak, Roman, Lechao Xiao, Jaehoon Lee, Yasaman Bahri, Greg Yang, Jiri Hron, Daniel A Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein (2018), “Bayesian deep convolutional networks with many channels are Gaussian processes,” *arXiv*, 1810.05148.
- Nyblom, Jukka (1989), “Testing for the constancy of parameters over time,” *Journal of the American Statistical Association*, 84, 223–230.
- Primiceri, Giorgio E. (2005), “Time varying structural vector autoregressions and monetary policy,” *The Review of Economic Studies*, 72, 821–852.
- Robinson, Peter M. (1991), “Time-varying nonlinear regression,” in *Economic Structural Change* eds. by Peter Hackl, and Anders Holger Westlund, 179–190, Berlin, Heidelberg: Springer Berlin Heidelberg.
- Stock, James H., and Mark W. Watson (1996), “Evidence on structural instability in macroeconomic time series relations,” *Journal of Business & Economic Statistics*, 14, 11–30.
- Stock, James H, and Mark W Watson (2007), “Why has US inflation become harder to forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.
- Stone, Charles J (1982), “Optimal global rates of convergence for nonparametric regression,” *The Annals of Statistics*, 1040–1053.
- Teckentrup, Aretha L (2020), “Convergence of gaussian process regression with estimated hyperparameters and applications in bayesian inverse problems,” *SIAM/ASA Journal on Uncertainty Quantification*, 8, 1310–1337.
- Teräsvirta, Timo (1994), “Specification, estimation, and evaluation of smooth transition autoregressive models,” *Journal of the American Statistical Association*, 89, 208–218.
- Tong, Howell (1990), *Non-linear Time Series: A Dynamical System Approach*: Oxford University Press.
- van der Vaart, Aad W, and J Harry van Zanten (2008), “Rates of contraction of posterior distributions based on Gaussian process priors,” *The Annals of Statistics*, 36, 1435–1463.
- Van Dyk, David A, and Taeyoung Park (2008), “Partially collapsed Gibbs samplers: Theory and methods,” *Journal of the American Statistical Association*, 103, 790–796.
- Williams, Christopher K., and Carl Edward Rasmussen (2006), *Gaussian processes for machine learning*, 2: MIT press Cambridge, MA.
- Yang, Yun, Anirban Bhattacharya, and Debdeep Pati (2017), “Frequentist coverage and sup-norm convergence rate in Gaussian process regression,” *arXiv*, 1708.04753.

Online Appendix

Gaussian Process Vector Autoregressions and Macroeconomic Uncertainty

Niko HAUZENBERGER^a, Florian HUBER^a, Massimiliano MARCELLINO^b, and Nico PETZ^a

^a*University of Salzburg*

^b*Bocconi University, IGER and CEPR*

Table of Contents

A	Technical Appendix	1
A.1	Capturing persistence through the kernel	1
A.2	Sampling the factors under linear restrictions	2
A.3	Sampling the log-volatilities	3
A.4	A sketch of the posterior simulator	7
A.5	Computing generalized impulse response functions	9
B	Data Appendix	11
C	Empirical Appendix	12
C.1	In-sample results for non-focus variables in the GP-VAR-8	12
C.2	Robustness with respect to different orderings	13
C.3	Impulse responses of non-focus variables in the GP-VAR-8	14
C.4	Additional results on model size and asymmetries	17

A Technical Appendix

A.1 Capturing persistence through the kernel

In this sub-section, we discuss how to handle persistent time series with GPs. In principle, appropriately choosing κ allows for capturing slowly evolving trends in y_t . But one elegant aspect of GPs is that stochastic trends in y_t can also be modeled explicitly through the kernel. This can be achieved as follows. Let \mathbf{B} denote a $T \times T$ lower triangular matrix with 1's on the main and off-diagonal elements.

The corresponding weight-space representation is then given by:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}_T, r\mathbf{I}_T), \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_T, \sigma_\varepsilon^2\mathbf{I}_T),$$

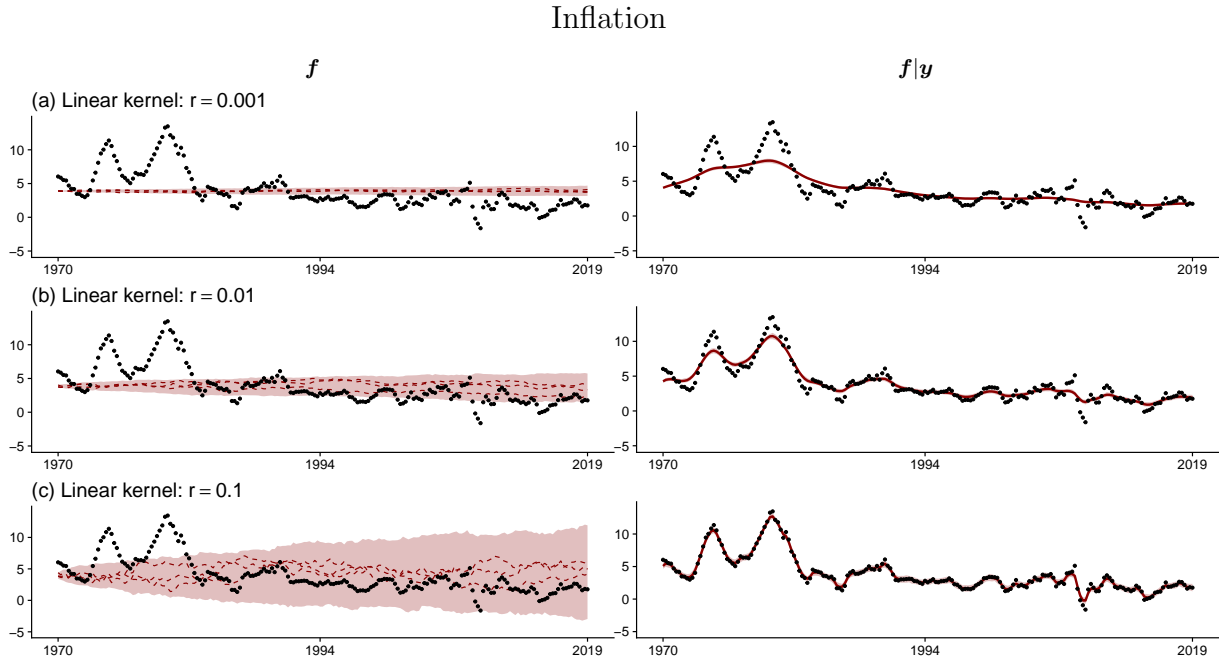
where r is a prior scaling parameter that controls the average jump size of the shocks in η_t . Notice that this equation can be represented in component form as follows:

$$y_t = \sum_{s=1}^t \eta_s + \varepsilon_t, \quad \eta_t \sim \mathcal{N}(0, r), \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2),$$

which implies that y_t is driven by a latent random walk factor with the matrix \mathbf{B} capturing the state evolution dynamics and r representing the state innovation variance. This is a standard unobserved components model. Notice that the corresponding kernel matrix can be derived as $r\mathbf{B}\mathbf{B}'$. Adding this to the Gaussian kernel discussed in the main text yields a combination between a kernel that captures unknown relations between y_t and \mathbf{x}_t but also possible stochastic trends in y_t .

The properties of this linear persistence kernel are illustrated in [Figure A.1](#). This figure again shows CPI inflation (again in year-on-year terms) but uses a longer sample (1970Q1 to 2019Q4) and the prior on \mathbf{f} (under the linear persistence kernel) and the posterior of \mathbf{f} . The figure reveals that if we set r close to zero, the corresponding estimate will only capture low frequency movements in y_t . The larger r gets, the larger the in-sample fit effectively becomes. If we set $r = 0.1$, we observe that the model yields an almost perfect in-sample fit.

Figure A.1: Effect of different values of r on the prior of \mathbf{f} and the posterior $\mathbf{f}|\mathbf{y}$



Notes: In this figure we showcase the GP regression with US inflation data, setting the matrix of regressors, \mathbf{B} , to be a lower triangular matrix with 1's as the diagonal and off-diagonal elements. The linear kernel is given by $r\mathbf{B}\mathbf{B}'$. The left panels report, for different values of r , the 5th and 95th prior percentiles (with the area in between shaded in light red), three draws from the prior (in dashed red), and the actual values of inflation (black dots). The right panels report the 90% posterior credible sets (shaded in light red), the posterior medians (in solid red), and actual inflation (black dots).

A.2 Sampling the factors under linear restrictions

As stated in the main text, we identify the conditional mean of the process by introducing the linear restriction that $\mathbf{G}\mathbf{g}_j = 0$ with $\mathbf{G} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. This can be efficiently achieved through the sampler proposed in Cong, Chen, and Zhou (2017).

Suppose our goal is to simulate \mathbf{g}_j from $\mathcal{N}_S(\bar{\mathbf{g}}_j, \bar{\mathbf{V}}_{\mathbf{g}_j})$ with restriction $\mathcal{S} = \{\mathbf{g}_j : \mathbf{G}\mathbf{g}_j = 0\}$ and moments given by:

$$\begin{aligned}\bar{\mathbf{V}}_{\mathbf{g}_j} &= \sqrt{\bar{\Omega}_j} \left(K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) - K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) (K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) + \mathbf{I}_T)^{-1} K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) \right) \sqrt{\bar{\Omega}_j}, \\ \bar{\mathbf{g}}_j &= \sqrt{\bar{\Omega}_j} K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) (K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) + \mathbf{I}_T)^{-1} \sqrt{\bar{\Omega}_j}^{-1} \left(\mathbf{Y}_j - \mathbf{f}_j - \sum_{k=1}^{j-1} q_{jk} \mathbf{Y}_k \right).\end{aligned}$$

This can be achieved using Algorithm 2 of Cong, Chen, and Zhou (2017). This algorithm consists of two steps. In the first step we sample \mathbf{g}_j^* from the unrestricted distribution $\mathcal{N}(\bar{\mathbf{g}}_j, \bar{\mathbf{V}}_{\mathbf{g}_j})$. In the second step, we obtain a draw from the restricted distribution by setting:

$$\mathbf{g}_j = \mathbf{g}_j^* - \bar{\mathbf{V}}_{\mathbf{g}_j} \mathbf{G}' (\mathbf{G} \bar{\mathbf{V}}_{\mathbf{g}_j} \mathbf{G}')^{-1} \mathbf{G} \mathbf{g}_j^*.$$

All involved quantities are trivial to compute and this step does not introduce additional computational hurdles. This normalization essentially subtracts a constant term from \mathbf{g}_j^* , the unrestricted draw of \mathbf{g}_j , so that the resulting grand mean will be equal to zero. This leaves the conditional mean function \mathbf{m}_j unchanged.

A.3 Sampling the log-volatilities

We sample the log-volatilities marginally of the factors. This is achieved by exploiting the weight-space view of the GP. Integrating out \mathbf{f}_j and \mathbf{g}_j allows us to rewrite the j^{th} equation as

$$\tilde{\mathbf{Y}}_j = \left(\mathbf{Y}_j - \sum_{k=1}^{j-1} q_{jk} \mathbf{Y}_k \right) = \sqrt{\tilde{\boldsymbol{\Omega}}_j} \tilde{\mathbf{W}}_j \tilde{\boldsymbol{\varepsilon}}_j, \quad \tilde{\boldsymbol{\varepsilon}}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T), \quad (\text{A.1})$$

where $\tilde{\mathbf{W}}_j$ denotes the lower Cholesky factor of $(K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + K_{\boldsymbol{\vartheta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j) + \mathbf{I}_T)$ and $\tilde{\boldsymbol{\Omega}}_j = \text{diag}(\boldsymbol{\omega}_j)$ with $\boldsymbol{\omega}_j = (\omega_{j1}, \dots, \omega_{jT})'$. Our goal is to sample the log-volatilities $\mathbf{h}_j = \log \boldsymbol{\omega}_j$ from its conditional posterior distribution $p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj})$, with $\boldsymbol{\theta}_{hj} = (\rho_{hj}, \sigma_{hj}^2, h_{j0})'$ collecting the parameters associated with the state equation of the log-volatility process, which evolves according to a stationary AR(1).

The corresponding T -dimensional full conditional posterior distribution can be expressed as:

$$p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj}) \propto p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j) \times p(\mathbf{h}_j | \boldsymbol{\theta}_{hj}), \quad (\text{A.2})$$

where $p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j)$ refers to the likelihood and $p(\mathbf{h}_j | \boldsymbol{\theta}_{hj})$ to the prior. While the prior is Gaussian and defined by the state equation in Eq. (3), the likelihood takes a multivariate Gaussian form only when conditioned on the log-volatilities \mathbf{h}_j (and thus $\tilde{\boldsymbol{\Omega}}_j$):

$$p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j) = (2\pi)^{-\frac{T}{2}} \times \det \left(\sqrt{\tilde{\boldsymbol{\Omega}}_j} \tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \sqrt{\tilde{\boldsymbol{\Omega}}_j} \right)^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \left(\tilde{\mathbf{Y}}_j' \left(\sqrt{\tilde{\boldsymbol{\Omega}}_j} \tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \sqrt{\tilde{\boldsymbol{\Omega}}_j} \right)^{-1} \tilde{\mathbf{Y}}_j \right) \right\}, \quad (\text{A.3})$$

where $\sqrt{\tilde{\boldsymbol{\Omega}}_j} \tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \sqrt{\tilde{\boldsymbol{\Omega}}_j}$ refers to the variance-covariance matrix of $\tilde{\mathbf{Y}}_j$. At this point it proves convenient to use the fact that $\tilde{\boldsymbol{\Omega}}_j$ is a diagonal matrix. The main challenge is that \mathbf{h}_j enters Eq. (A.3) non-linearly and is never observed directly (but evolves according to

a latent process), which complicates the evaluation of the likelihood.

Moreover, the conjugate nature of our GP regression complicates likelihood evaluation further. The algorithm proposed by Kim, Shephard, and Chib (1998) based on auxiliary mixture indicators cannot be used because $\boldsymbol{\Omega}_j$ – and thus \mathbf{h}_j – are serially correlated and are not independent from each other, as the full symmetric kernel matrices $K_{\boldsymbol{\theta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j)$ and $K_{\boldsymbol{\theta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j)$ enter the Cholesky factor $\tilde{\mathbf{W}}_j$.¹

As a remedy, we follow Chan (2017) who proposes computationally efficient sampling techniques. Our algorithm is a variant of the independence Metropolis-Hastings (MH) which can be readily applied in the present setting.² As noted by Chan (2017), this algorithm is fast and exhibits good mixing properties for several reasons. The independent MH step samples the log-volatilities jointly from their full conditional posterior distribution, the proposal distribution is constructed such that the acceptance rate of the MH step is sufficiently high, and in addition, we resort to sparse matrix methods to speed up computation.

In the following, we provide a detailed discussion on our independence MH step. We use a second-order Taylor approximation with respect to \mathbf{h}_j and then employ the Newton-Raphson optimization algorithm to represent the non-trivial conditional posterior in Eq. (A.2) in the form of a multivariate Gaussian distribution. This approximation is centered on the mode of $p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj})$ and the variance-covariance is given by the inverse of the negative Hessian of $\log p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj})$ evaluated at the obtained mode. This Gaussian distribution provides a sufficiently accurate approximation to the full conditional posterior of \mathbf{h}_j and thus ensures a high acceptance rate when used as the proposal density in our MH step, leading to favorable mixing properties.

To construct our Gaussian proposal density thus requires evaluating the gradient and Hessian of the logarithm of the full conditional posterior of \mathbf{h}_j . According to Eq. (A.2), the log-posterior distribution is the sum of the log-likelihood and log-prior distribution, implying that the gradient/Hessian of the log-posterior is also the sum of the gradient/Hessian of these two components.

We first focus on the gradient and the Hessian of the log-prior distribution, since

¹In particular, this implies that the measurement errors $\log(\tilde{\boldsymbol{\epsilon}}_j^2)$ of the linearized observation equation (which is obtained by squaring and taking the logarithm of Eq. (A.1)) are no longer independently $\log(\chi_1^2)$ distributed and cannot readily be approximated by a known mixture of Gaussian distribution to render the likelihood Gaussian conditional on the mixture indicators.

²For a textbook treatment, see Chan, et al. (2019).

these quantities are mostly standard (see, e.g., Chan, et al., 2019). For the t^{th} element of \mathbf{h}_j , the prior is defined by Eq. (3) and depends on the additional state equation hyperparameters $\boldsymbol{\theta}_{hj}$. For \mathbf{h}_j , Eq. (3) can be written more compactly as:

$$\begin{aligned} \mathbf{D}_{hj}\mathbf{h}_j &= \tilde{\boldsymbol{\mu}}_{hj} + \tilde{\mathbf{u}}_{hj}, & \tilde{\mathbf{u}}_{hj} &\sim \mathcal{N}(\mathbf{0}_T, \sigma_{hj}^2 \mathbf{I}_T), \\ \mathbf{h}_j &= \boldsymbol{\mu}_{hj} + \mathbf{u}_{hj}, & \mathbf{u}_{hj} &\sim \mathcal{N}(\mathbf{0}_T, \sigma_{hj}^2 (\mathbf{D}'_{hj}\mathbf{D}_{hj})^{-1}), \end{aligned} \quad (\text{A.4})$$

with

$$\mathbf{D}_{hj} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -\rho_{hj} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho_{hj} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho_{hj} & 1 \end{pmatrix}$$

denoting a $T \times T$ -matrix, $\tilde{\boldsymbol{\mu}}_{hj} = (\rho_{hj}h_{j0}, 0, \dots, 0)'$ being $T \times 1$ -vector, and $\boldsymbol{\mu}_{hj} = \mathbf{D}_{hj}^{-1}\tilde{\boldsymbol{\mu}}_{hj}$. Then the log-prior distribution of \mathbf{h}_j is given by:

$$\log p(\mathbf{h}_j|\boldsymbol{\theta}_{hj}) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_{hj}^2 - \frac{1}{2\sigma_{hj}^2} (\mathbf{h}_j - \boldsymbol{\mu}_{hj})' \mathbf{D}'_{hj} \mathbf{D}_{hj} (\mathbf{h}_j - \boldsymbol{\mu}_{hj}), \quad (\text{A.5})$$

by noting that $\det(\mathbf{D}'_{hj}\mathbf{D}_{hj}) = 1$.

In the following, let $\mathbf{r}_P(\tilde{\mathbf{h}}_j)$ denote the first derivative (gradient) and $\mathbf{R}_P(\tilde{\mathbf{h}}_j)$ the second derivative (Hessian) of the log-prior distribution with respect to \mathbf{h}_j evaluated at $\tilde{\mathbf{h}}_j$:

$$\begin{aligned} \mathbf{r}_P(\tilde{\mathbf{h}}_j) &= \left. \frac{\partial \log p(\mathbf{h}_j|\boldsymbol{\theta}_{hj})}{\partial \mathbf{h}_j} \right|_{\mathbf{h}_j=\tilde{\mathbf{h}}_j} = -\frac{1}{2\sigma_{hj}^2} \mathbf{D}'_{hj} \mathbf{D}_{hj} (\mathbf{h}_j - \boldsymbol{\mu}_{hj}), \\ \mathbf{R}_P(\tilde{\mathbf{h}}_j) &= \left. -\frac{\partial^2 \log p(\mathbf{h}_j|\boldsymbol{\theta}_{hj})}{\partial \mathbf{h}_j \partial \mathbf{h}'_j} \right|_{\mathbf{h}_j=\tilde{\mathbf{h}}_j} = -\frac{1}{2\sigma_{hj}^2} \mathbf{D}'_{hj} \mathbf{D}_{hj}. \end{aligned}$$

While $\mathbf{r}_P(\tilde{\mathbf{h}}_j)$ is a $T \times 1$ vector, $\mathbf{R}_P(\tilde{\mathbf{h}}_j)$ is a tridiagonal matrix of dimension T , with non-zero elements on the main diagonal and the two diagonals below and above the main one.

Next, we focus on the log-likelihood, which has the following form:

$$\begin{aligned} \log p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j) &= -\frac{T}{2} \log 2\pi - \frac{1}{2} \left(\sum_{t=1}^T h_{jt} \right) - \frac{1}{2} \log \left(\det \left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right) \right) \\ &\quad - \frac{1}{2} \left(\hat{\mathbf{Y}}_j' \left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1} \hat{\mathbf{Y}}_j \right), \end{aligned} \quad (\text{A.6})$$

where all the quantities involving $\tilde{\mathbf{W}}_j$ can be pre-computed and $\hat{\mathbf{Y}}_j = \left(\tilde{\mathbf{Y}}_j \odot \exp \left\{ -\frac{h_j}{2} \right\} \right)$ with \odot denoting the component-wise multiplication. Now, let a $T \times 1$ vector $\mathbf{r}_L(\tilde{\mathbf{h}}_j)$ denote the gradient and a $T \times T$ -matrix $\mathbf{R}_L(\tilde{\mathbf{h}}_j)$, the Hessian of the log-likelihood again with respect to \mathbf{h}_j evaluated at $\tilde{\mathbf{h}}_j$:

$$\begin{aligned} \mathbf{r}_L(\tilde{\mathbf{h}}_j) &= \left. \frac{\partial \log p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j)}{\partial \mathbf{h}_j} \right|_{\mathbf{h}_j = \tilde{\mathbf{h}}_j} = -\frac{1}{2} \mathbf{1}_T + \frac{1}{2} \left(\left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1} \hat{\mathbf{Y}}_j \right) \odot \hat{\mathbf{Y}}_j, \\ \mathbf{R}_L(\tilde{\mathbf{h}}_j) &= \left. -\frac{\partial^2 \log p(\tilde{\mathbf{Y}}_j | \mathbf{h}_j, \tilde{\mathbf{W}}_j)}{\partial \mathbf{h}_j \partial \mathbf{h}_j'} \right|_{\mathbf{h}_j = \tilde{\mathbf{h}}_j} = \frac{1}{4} \text{diag} \left(\hat{\mathbf{Y}}_j \right) \left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1} \text{diag} \left(\hat{\mathbf{Y}}_j \right) \\ &\quad + \frac{1}{4} \text{diag} \left(\left(\left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1} \hat{\mathbf{Y}}_j \right) \odot \hat{\mathbf{Y}}_j \right). \end{aligned}$$

It is easy to see that the presence of $\left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1}$ might lead to a fairly dense $T \times T$ Hessian, which makes the use of band sparse matrix algorithms impractical and leads to non-negligible computational costs for matrix calculations involving the Hessian.³

In what follows, we exploit the special structure of $\left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1}$. Elements of the two kernels $K_{\vartheta_{j1}}(\mathbf{X}_j, \mathbf{X}_j)$ and $K_{\vartheta_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j)$ are defined by a squared exponential function that per construction imposes a view that periods close to each other (usually this is the case for $t-1$, t , and $t+1$) feature relatively large covariances, while dissimilar periods have very small covariances. This structure implies that the elements of $\left(\tilde{\mathbf{W}}_j \tilde{\mathbf{W}}_j' \right)^{-1}$ on the tridiagonal main band carry most of the information, while entries far away from this main band are typically close to zero. Hence, we approximate $\mathbf{R}_L(\tilde{\mathbf{h}}_j)$ by a $T \times T$ matrix $\hat{\mathbf{R}}_L(\tilde{\mathbf{h}}_j)$ that keeps only the elements of $\mathbf{R}_L(\tilde{\mathbf{h}}_j)$ on the tridiagonal main band and sets all other entries to zero. This approximation has the convenient feature that both $\hat{\mathbf{R}}_L(\tilde{\mathbf{h}}_j)$ and $\mathbf{R}_P(\tilde{\mathbf{h}}_j)$ have the same band sparse structure. Consequently, the Hessian of the log-posterior distribution is also a tridiagonal matrix, which greatly facilitates computation through the use of band sparse matrix algorithms.

³A band sparse matrix is symmetric and contains only non-zero entries along the main diagonal band.

Combining the gradient (the approximated Hessian) of the log-likelihood with that of the log-prior distribution, we obtain the gradient (the approximated Hessian) of the log-posterior distribution for \mathbf{h}_j evaluated at $\tilde{\mathbf{h}}_j$:

$$\begin{aligned} \mathbf{r}(\tilde{\mathbf{h}}_j) &= \left. \frac{\partial \log p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj})}{\partial \mathbf{h}_j} \right|_{\mathbf{h}_j = \tilde{\mathbf{h}}_j} = \mathbf{r}_L(\tilde{\mathbf{h}}_j) + \mathbf{r}_P(\tilde{\mathbf{h}}_j), \\ \mathbf{R}(\tilde{\mathbf{h}}_j) &= - \left. \frac{\partial^2 \log p(\mathbf{h}_j | \tilde{\mathbf{Y}}_j, \tilde{\mathbf{W}}_j, \boldsymbol{\theta}_{hj})}{\partial \mathbf{h}_j \partial \mathbf{h}'_j} \right|_{\mathbf{h}_j = \tilde{\mathbf{h}}_j} = \mathbf{R}_L(\tilde{\mathbf{h}}_j) + \mathbf{R}_P(\tilde{\mathbf{h}}_j) \quad \text{and} \\ \mathbf{R}(\tilde{\mathbf{h}}_j) &\approx \hat{\mathbf{R}}(\tilde{\mathbf{h}}_j) = \hat{\mathbf{R}}_L(\tilde{\mathbf{h}}_j) + \mathbf{R}_P(\tilde{\mathbf{h}}_j). \end{aligned}$$

Finally, to define the moments of our Gaussian proposal density for our independence MH step, we employ the Newton-Raphson optimization method. To obtain the mode $\hat{\mathbf{h}}_j$ of the log-posterior distribution of \mathbf{h}_j , we initialize the algorithm with a suitable starting value $\mathbf{h}_j = \tilde{\mathbf{h}}_j^{(1)}$ and iterate through

$$\tilde{\mathbf{h}}_j^{(s+1)} = \tilde{\mathbf{h}}_j^{(s)} - \left(\hat{\mathbf{R}}(\tilde{\mathbf{h}}_j^{(s)}) \right)^{-1} \mathbf{r}(\tilde{\mathbf{h}}_j^{(s)})$$

until $\|\tilde{\mathbf{h}}_j^{(s+1)} - \tilde{\mathbf{h}}_j^{(s)}\| < \epsilon$, with $\epsilon = 10^{-4}$ being sufficiently close to zero and acting as convergence criterion for the numerical maximization algorithm. While $\hat{\mathbf{h}}_j$ is used as the mean of the Gaussian proposal density, the variance-covariance is set to the inverse of the (approximated) negative Hessian evaluated at $\hat{\mathbf{h}}_j$. Hence, the proposal is given by: $\mathcal{N}\left(\hat{\mathbf{h}}_j, \left(-\hat{\mathbf{R}}(\hat{\mathbf{h}}_j)\right)^{-1}\right)$.

This proposal has good empirical properties. In all our empirical work it leads to acceptance rates between 35 and 50 percent.

A.4 A sketch of the posterior simulator

Our posterior simulator is comprised of several steps that involve the full conditional distributions of the corresponding latent quantities and parameters of the model. Since we use a collapsed sampler the ordering of the steps is crucial for sampling from the correct joint posterior distribution.

The joint posterior distribution associated with the j^{th} equation is given by:

$$p(\mathbf{f}_j, \mathbf{g}_j, q_{j1}, \dots, q_{jj-1}, \boldsymbol{\Omega}_j, \boldsymbol{\vartheta}_{j1}, \boldsymbol{\vartheta}_{j2}, \mathbf{d}_j, \boldsymbol{\nu}_j, \boldsymbol{\varpi}_j | \text{Data}), \quad (\text{A.7})$$

where d_j denotes the mixture indicators used in the Gaussian approximation to the $\log\text{-}\chi_1^2$ distribution, $\boldsymbol{\nu}_j$ are the coefficients associated with the SV state equation and $\boldsymbol{\varpi}_j$ are the hyperparameters associated with the Horseshoe prior. Let $\mathbf{q}_{j\bullet} = (q_{j1}, \dots, q_{jj-1})'$ denote the j^{th} row of \mathbf{Q} .

Conditional on adequately chosen starting values, our MCMC algorithm draws from the joint posterior of equation j^{th} coefficients and states by iterating through the following steps.

- Step 1: Sample a value of $\mathbf{q}_{j\bullet}$ from $p(\mathbf{q}_{j\bullet} | \mathbf{f}_j, \mathbf{g}_j, \boldsymbol{\Omega}_j, \boldsymbol{\varpi}_j, \text{Data}) \sim \mathcal{N}(\bar{\mathbf{q}}_{j\bullet}, \bar{\mathbf{V}}_{q_j})$ if $j > 1$.⁴ The moments $\bar{\mathbf{V}}_{q_j}$ and $\bar{\mathbf{q}}_{j\bullet}$ take a standard form.
- Step 2: Simulate the full history of the (log) volatilities from $p(\boldsymbol{\Omega}_j | \mathbf{q}_{j\bullet}, \boldsymbol{\vartheta}_{j1}, \boldsymbol{\vartheta}_{j2}, \mathbf{d}_j, \boldsymbol{\nu}_j, \text{Data})$ using the independent MH algorithm outlined in Sub-section A.3. Notice that this step is marginally of \mathbf{f}_j and \mathbf{g}_j
- Step 3: The factor \mathbf{f}_j is obtained from the multivariate Gaussian conditional posterior $p(\mathbf{f}_j | \mathbf{g}_j, \mathbf{q}_{j\bullet}, \boldsymbol{\vartheta}_{j1}, \boldsymbol{\Omega}_j, \text{Data})$ as described in Sub-section 3.4.
- Step 4: The factor \mathbf{g}_j is obtained from the multivariate Gaussian conditional posterior $p(\mathbf{g}_j | \mathbf{f}_j, \mathbf{q}_{j\bullet}, \boldsymbol{\vartheta}_{j2}, \boldsymbol{\Omega}_j, \text{Data})$ as described in Sub-section A.2.
- Step 5: $\boldsymbol{\vartheta}_{j1}$ is obtained from its discrete posterior distribution using multinomial sampling.
- Step 6: Similarly, $\boldsymbol{\vartheta}_{j2}$ is also obtained through multinomial sampling.
- Step 7: Sample the parameters of state equation associated with the log-volatilities. The corresponding conditional distributions all take standard forms.
- Step 8: Sample the hyperparameters of the Horseshoe prior in $\boldsymbol{\varpi}$. We use the Gibbs updating step described in Makalic and Schmidt (2015) which only samples from inverse Gamma distributions.

The precise ordering of the steps in our algorithm is crucial for obtaining draws from the correct stationary distribution since we rely on marginalizing out some of the states/parameters in some steps of the MCMC algorithm.⁵ Notice that \mathbf{f}_j and \mathbf{g}_j are sampled conditionally

⁴If $j = 1$, ignore this step and proceed to step 2.

⁵Van Dyk and Park (2008) discuss a general approach on how to construct collapsed Gibbs samplers in order to maintain the correct stationary distribution.

on $\boldsymbol{\Omega}_j$ while $\boldsymbol{\Omega}_j$ is sampled marginally from \boldsymbol{f}_j and \boldsymbol{g}_j . This implies that we again sample $\boldsymbol{\Omega}_j$ marginally from the factors and the factors conditionally on the volatilities, leading to a joint update from $p(\boldsymbol{\Omega}_j, \boldsymbol{f}_j, \boldsymbol{g}_j | \bullet)$.

A.5 Computing generalized impulse response functions

Computing the GIRFs is achieved as follows. We proceed in an equation-by-equation basis using the structural representation of the GP-VAR in Eq. (2). In this case, the one-step-ahead predictive distribution of m_{1t} is:

$$m_{1t+1} \sim \mathcal{N}(\bar{m}_{1t+1}, \bar{V}_{1t+1}),$$

with predictive moments given by:

$$\begin{aligned} \bar{V}_{1t+1} &= \sigma_{1t+1} \left(k_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_{1t+1}, \mathbf{W}_{1t+1}) - K_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_{1t+1}, \mathbf{W}_1)(K_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_1, \mathbf{W}_1) + \mathbf{I}_T)^{-1}K_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_1, \mathbf{W}_{1t+1}) \right), \\ \bar{m}_{1t+1} &= \sigma_{1t+1} K_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_{1t+1}, \mathbf{W}_1)(K_{\boldsymbol{\vartheta}_1}^*(\mathbf{W}_1, \mathbf{W}_1) + \mathbf{I}_T)^{-1} \sqrt{\overline{\boldsymbol{\Omega}}_1}^{-1} \mathbf{Y}_1. \end{aligned}$$

Here, we let $\mathbf{W}_j = (\mathbf{X}_j, \mathbf{Z}_j)$ with t^{th} row \mathbf{W}_{jt} and $K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_j, \mathbf{W}_j) = K_{\boldsymbol{\vartheta}_{j1}}(\mathbf{X}_j, \mathbf{X}_j) + K_{\boldsymbol{\vartheta}_{j2}}(\mathbf{Z}_j, \mathbf{Z}_j)$. A draw from the one-step-ahead predictive distribution of y_{1t+1} is obtained from:

$$y_{1t+1} \sim \mathcal{N}(m_{1t+1}, \omega_{1t}).$$

Draws from the forecast distribution of y_{1t+1} are used to form the predictive distribution for m_{2t} (and thus y_{2t+1} and so on). In general, the predictive density for the conditional mean of equation $j > 2$ is given by:

$$m_{jt+1} \sim \mathcal{N}(\bar{m}_{jt+1}, \bar{V}_{jt+1}),$$

with

$$\bar{m}_{jt+1} = \sigma_{jt+1} K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_{jt+1}, \mathbf{W}_j)(K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_j, \mathbf{W}_j) + \mathbf{I}_T)^{-1} \sqrt{\overline{\boldsymbol{\Omega}}_j}^{-1} \left(\mathbf{Y}_j - \sum_{k=1}^{j-1} q_{jk}(m_{kt+1} + \varepsilon_{kt+1}) \right),$$

and the predictive variance defined as in the case of $j = 1$. A draw from the predictive

distribution of y_{jt+1} is obtained by drawing $\varepsilon_{jt+1} \sim \mathcal{N}(0, \omega_{jt+1})$ and adding this to m_{jt+1} .

Drawing from the posterior of the forecasts for all M elements in \mathbf{y}_t yields the one-step-ahead forecast distribution $p(\mathbf{y}_{t+1}|\mathcal{I}_t)$. Higher order forecast distributions are then obtained by using $\mathbf{y}_{t+1} \sim p(\mathbf{y}_{t+1}|\mathcal{I}_t)$ to construct $\overline{\mathbf{W}}_{jt+2}$ and then compute \overline{m}_{jt+2} , for $j = 1, \dots, M$, and drawing from the marginal distribution of the structural shocks. In general, h -step-ahead predictions \mathbf{y}_{t+h} are obtained similarly by drawing from $p(\mathbf{y}_{t+h}|\mathcal{I}_t)$.

As described in the text, we focus on an uncertainty shock with the uncertainty index being located in the j^{th} position in \mathbf{y}_t . Let ε_{jt} denote the structural innovation in time t and we assume that $\varepsilon_{jt} = \varsigma$. This implies that \mathbf{y}_t changes by \mathbf{q}_j . Again, we can compute point forecasts but instead of using \mathbf{W}_{jt+1} (which comprises of \mathbf{y}_t and its $p-1$ lags), $\hat{\mathbf{W}}_{jt+1}$ is constructed based on $\hat{\mathbf{y}}_t = \mathbf{y}_t + \mathbf{q}_1 + \tilde{\boldsymbol{\varepsilon}}_t$. The shock $\tilde{\boldsymbol{\varepsilon}}_t$ is obtained from the marginal distribution of the shocks but its j^{th} element equals zero. Using $\hat{\mathbf{W}}_{jt+1}$, we can compute the predictive distributions of \mathbf{y}_{t+1} given a shock of size ς in ε_{jt} . As an intermediate step, we need to compute the mean forecast based on $\varepsilon_{jt} = \varsigma$:

$$\hat{m}_{jt+1} \sim \mathcal{N}(\overline{m}_{jt+1}^\dagger, \overline{V}_{jt+1}^\dagger)$$

with mean and variance given by:

$$\begin{aligned} \overline{m}_{jt+1}^\dagger &= \sigma_{jt+1} K_{\boldsymbol{\vartheta}_j}^*(\hat{\mathbf{W}}_{jt+1}, \mathbf{W}_j) (K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_j, \mathbf{W}_j) + \mathbf{I}_T)^{-1} \sqrt{\boldsymbol{\Omega}_j}^{-1} (\mathbf{Y}_j - \sum_{k=1}^{j-1} q_{jk} (\hat{m}_{kt+1} + \varepsilon_{kt+1})), \\ \overline{V}_{jt+1}^\dagger &= \sigma_{jt+1} \left(k_{\boldsymbol{\vartheta}_j}^*(\hat{\mathbf{W}}_{1t+1}, \hat{\mathbf{W}}_{1t+1}) - K_{\boldsymbol{\vartheta}_j}^*(\hat{\mathbf{W}}_{1t+1}, \mathbf{W}_j) (K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_j, \mathbf{W}_j) + \mathbf{I}_T)^{-1} K_{\boldsymbol{\vartheta}_j}^*(\mathbf{W}_j, \hat{\mathbf{W}}_{1t+1}) \right) \end{aligned}$$

These one-step-ahead forecasts can again be used to construct iterative higher order predictions conditional on a unit structural shock to the first variable. Doing so for each of the M variables yields a draw from $y_{t+h} \sim p(\mathbf{y}_{t+h}|\mathcal{I}_t, \varepsilon_{jt} = \varsigma)$.

B Data Appendix

Table B.1: Data description.

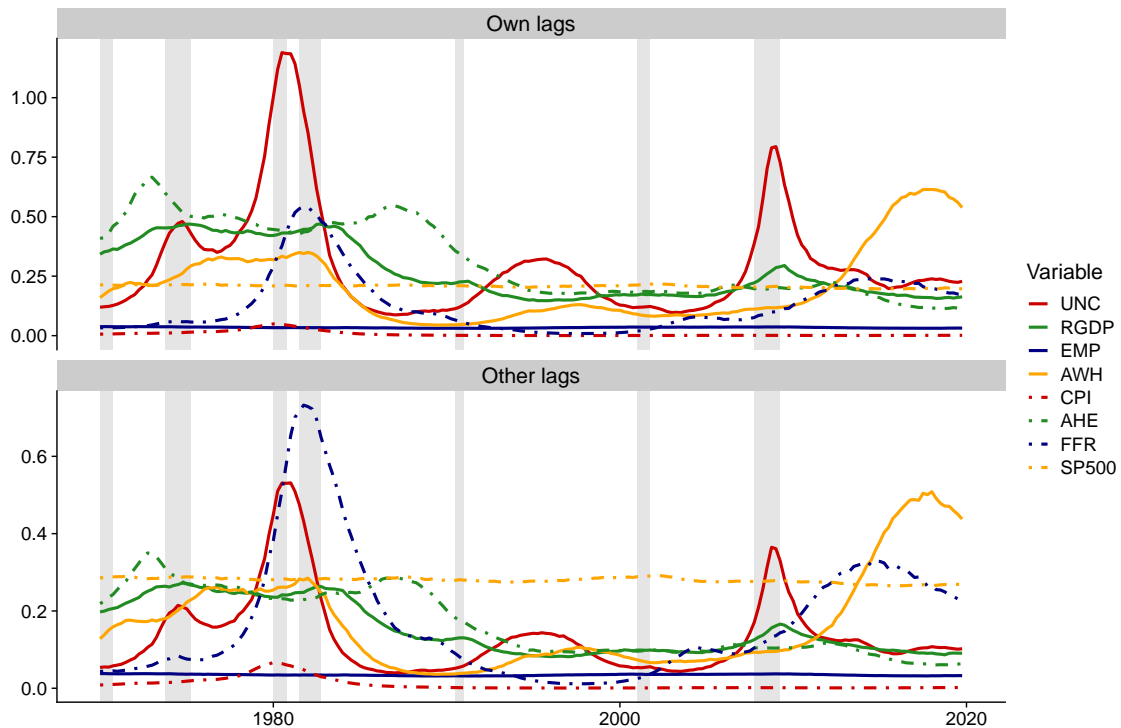
Mnemonic	Description	Trans.	VAR-8	VAR-16	VAR-32	VAR-64
UNC	Macroeconomic Uncertainty Index of Jurado, Ludvigson, and Ng (2015)	1	x	x	x	x
GDPC1 (RGDP)	Real Gross Domestic Product	2	x	x	x	x
CE16OV (EMP)	Civilian Employment (Thousands of Persons)	2	x	x	x	x
AWHMAN (AWH)	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	1	x	x	x	x
CPIAUCSL (CPI)	Consumer Price Index for All Urban Consumers: All Items	2	x	x	x	x
CES3000000008x (AHE)	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	2	x	x	x	x
FEDFUNDS (FFR)	Effective Federal Funds Rate (Percent)	1	x	x	x	x
S.P.500 (SP500)	S&P's Common Stock Price Index: Composite	3	x	x	x	x
PCECC96	Real Personal Consumption Expenditures	2		x	x	x
FPIx	Real private fixed investment	2		x	x	x
UNRATE	Civilian Unemployment Rate (Percent)	1		x	x	x
CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	1		x	x	x
CLAIMSx	Initial Claims	2		x	x	x
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	2		x	x	x
CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees:	2		x	x	x
M2REAL	Real M2 Money Stock	2		x	x	x
GCEC1	Real Government Consumption Expenditures and Gross Investment	2			x	x
INDPRO	IP:Total index Industrial Production Index (Index 2012=100)	2			x	x
CUMFNS	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)	1			x	x
PAYEMS	Emp:Nonfarm All Employees: Total nonfarm (Thousands of Persons)	2			x	x
PERMIT	New Private Housing Units Authorized by Building Permits	2			x	x
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	2			x	x
GDPCPTPI	Gross Domestic Product: Chain-type Price Index	2			x	x
CES2000000008x	Real Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	2			x	x
BAA10YM	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury	1			x	x
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market	1			x	x
TB3SMFPM	3-Month Treasury Constant Maturity Minus Federal Funds Rate	1			x	x
AAAFEM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	1			x	x
BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks	2			x	x
CONSUMERx	Real Consumer Loans at All Commercial Banks	2			x	x
NONREVSx	Total Real Nonrevolving Credit Owned and Securitized, Outstanding	2			x	x
NONBORRES	Reserves Of Depository Institutions, Nonborrowed	4			x	x
GPDI1	Real Gross Private Domestic Investment	2				x
PNFIx	Real private fixed investment: Nonresidential	2				x
PRFIx	Real private fixed investment: Residential	2				x
EXPGSC1	Real Exports of Goods and Services	2				x
IMPGSC1	Real Imports of Goods and Services	2				x
IPCONGD	IP:Consumer goods Industrial Production: Consumer Goods (Index 2012=100)	2				x
UNRATELTx	Unemployment Rate for more than 27 weeks (Percent)	1				x
AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	1				x
AMDMNOx	Real Manufacturers' New Orders: Durable Goods (Millions of 2012 Dollars)	2				x
GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index	2				x
DGDSRG3Q086SBEA	Personal consumption expenditures: Goods	2				x
DDURRG3Q086SBEA	Personal consumption expenditures: Durable goods	2				x
DSERRG3Q086SBEA	Personal consumption expenditures: Services	2				x
DNDGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods	2				x
CPILFESL	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	2				x
OILPRICEx	Real Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma	2				x
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour (Index 2012=100)	2				x
RCPHBS	Business Sector: Real Compensation Per Hour (Index 2012=100)	2				x
TB3MS	3-Month Treasury Bill: Secondary Market Rate (Percent)	1				x
TB6MS	6-Month Treasury Bill: Secondary Market Rate (Percent)	1				x
GS1	1-Year Treasury Constant Maturity Rate (Percent)	1				x
GS10	10-Year Treasury Constant Maturity Rate (Percent)	1				x
AAA	Moody's Seasoned Aaa Corporate Bond Yield (Percent)	1				x
BAA	Moody's Seasoned Baa Corporate Bond Yield (Percent)	1				x
TB6M3Mx	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)	1				x
GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market	1				x
CF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market	1				x
MIREAL	Real M1 Money Stock	2				x
REALLNx	Real Real Estate Loans, All Commercial Banks	2				x
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	2				x
S.P.indust	S&P's Common Stock Price Index: Industrials	3				x
S.P.div.yield	S&P's Composite Common Stock: Dividend Yield	1				x

Notes: We use the macroeconomic uncertainty measure of Jurado, Ludvigson, and Ng (2015) provided (and regularly updated) on the web page of Sydney C. Ludvigson (available online via sydneyludvigson.com/macro-and-financial-uncertainty-indexes). Otherwise, we rely on the quarterly version of the dataset proposed in McCracken and Ng (2016). **Trans** indicates the transformation applied to each variable with (1) implying no transformation, (2) denoting year-on-year growth rates, (3) denoting quarter-on-quarter growth rates, and (4) refers to quarter-on-quarter percentage changes.

C Empirical Appendix

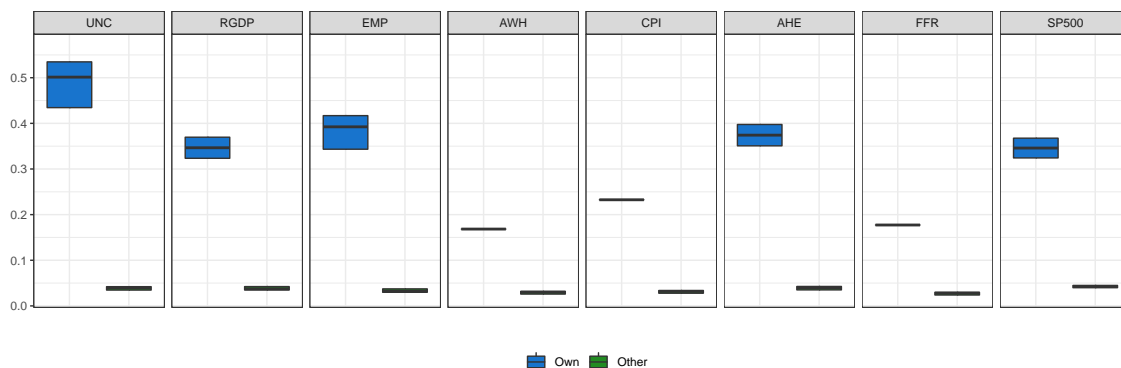
C.1 In-sample results for non-focus variables in the GP-VAR-8

Figure C.1: Linear shrinkage parameters of equation-specific kernels for the GP-VAR-64.



Notes: This figure reports the posterior means of the product of the error variances ω_{jt} and the linear scaling parameters for own lags (ξ_{j1}) and for other lags (ξ_{j2}), respectively. These two quantities correspond to the diagonal elements of the re-scaled kernels $\omega_{jt} \times k_{\vartheta_{j1}}(\mathbf{x}_t, \mathbf{x}_t) = \omega_{jt}\xi_{j1}$ and $\omega_{jt} \times k_{\vartheta_{j2}}(\mathbf{z}_t, \mathbf{z}_t) = \omega_{jt}\xi_{j2}$.

Figure C.2: Inverse length scale parameters of equation-specific kernels for the GP-VAR-64.

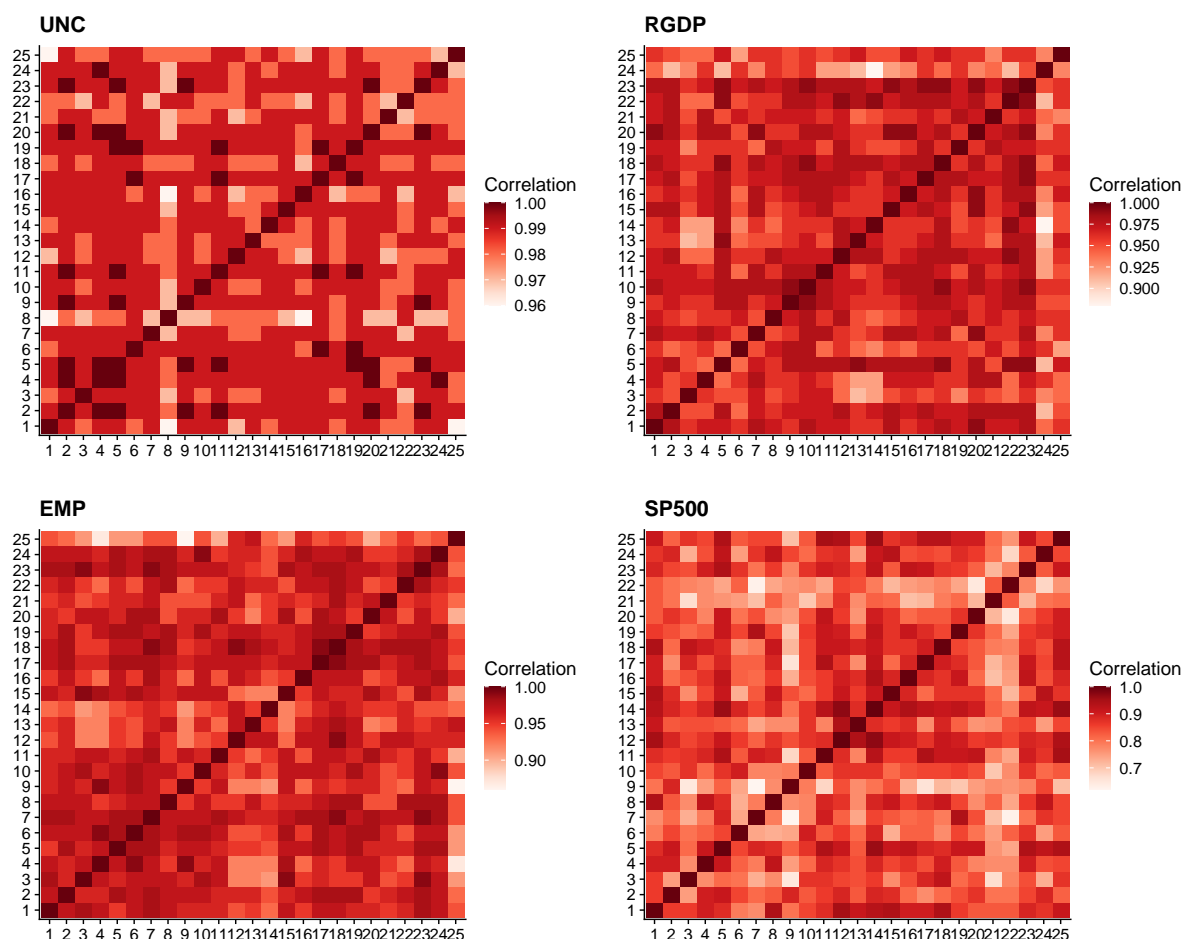


Notes: This figure reports the posterior summaries in the form of simplified boxplots of the inverse length scale parameters for own lags (κ_{j1}) and other lags (κ_{j2}), respectively. The solid black lines denote the posterior medians, while the blue (green) shaded areas represent the 50% posterior credible sets (i.e., the posterior interquartile ranges).

C.2 Robustness with respect to different orderings

This figure shows heatmaps of the correlations between the posterior median of the GIRFs for 25 different variable orderings (chosen at random). To ensure that the identification does not impact the results, we fix the ordering of the first two elements in \mathbf{y}_t (the S&P 500 is ordered first and the uncertainty index second). The heatmaps show that for our focus variables, the correlations are (almost) always above 0.9. This indicates that the ordering does not play a particular role for the estimation of the IRFs. The only responses that display slightly stronger changes are the ones of the S&P 500. In this case, the differences in IRFs are, however, mostly related to higher order IRFs which are insignificant. Short-term reactions are almost identical and thus do not change much if we alter the orderings.

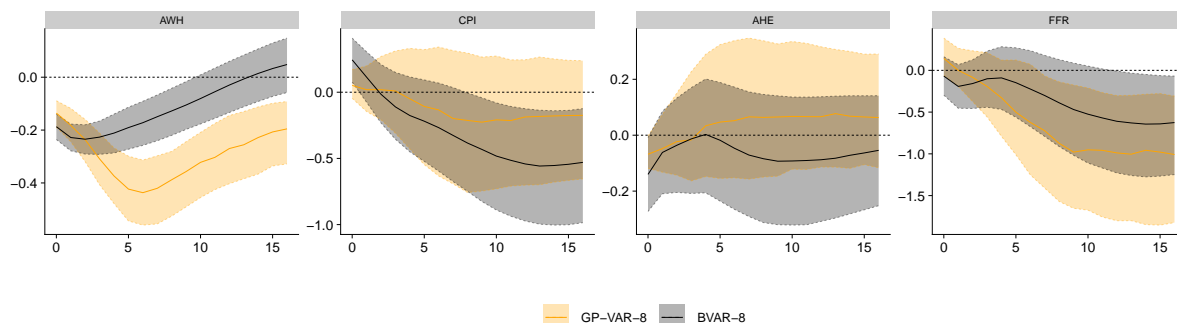
Figure C.3: Correlation of GIRFs across different orderings in the estimation of the GP-VAR-8.



Notes: The figure shows, for each focus variable, the correlation of the median of the average GIRF for 25 different variable orderings in the estimation. The macroeconomic uncertainty index is always ordered second and the S&P 500 is always ordered first. The reference model specification used is the GP-VAR-8.

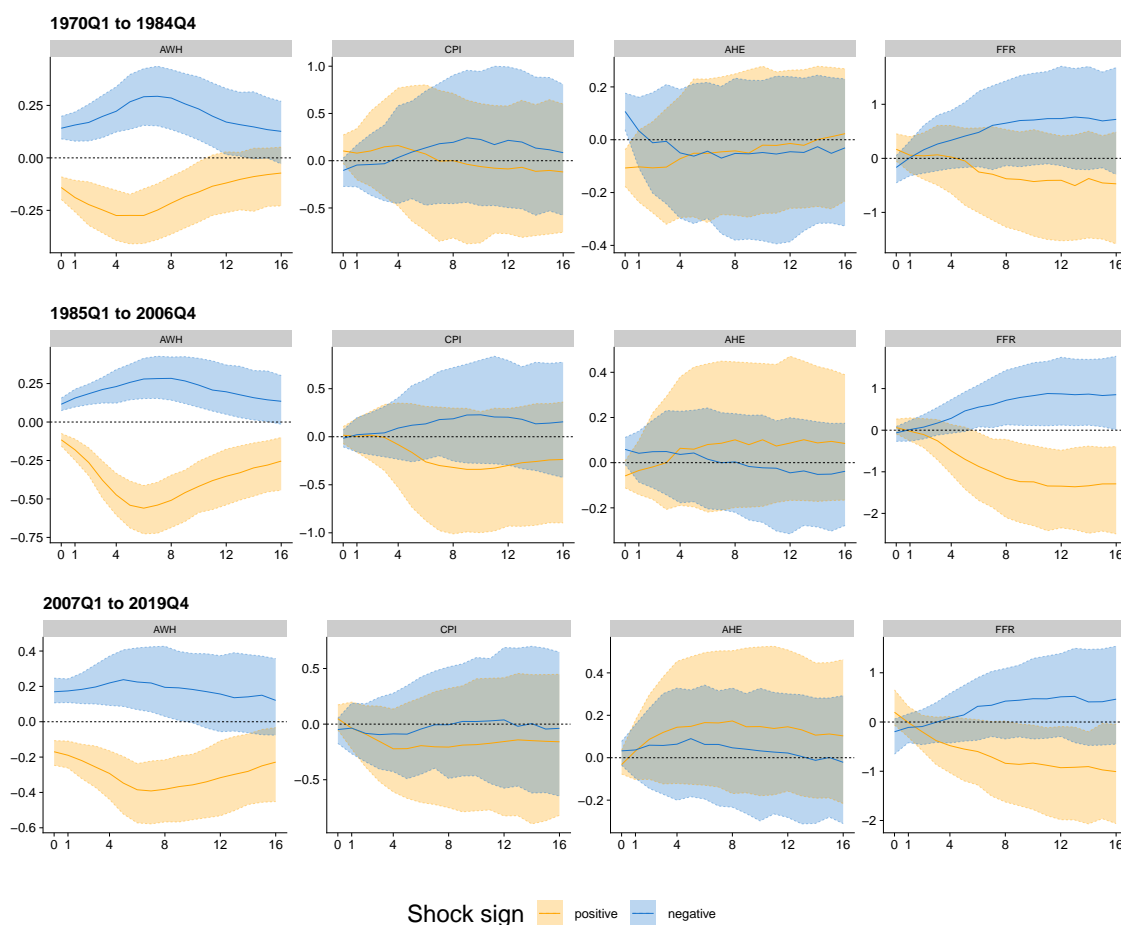
C.3 Impulse responses of non-focus variables in the GP-VAR-8

Figure C.4: Impulse responses of non-focus variables in the GP-VAR-8 relative to a small-scale BVAR.



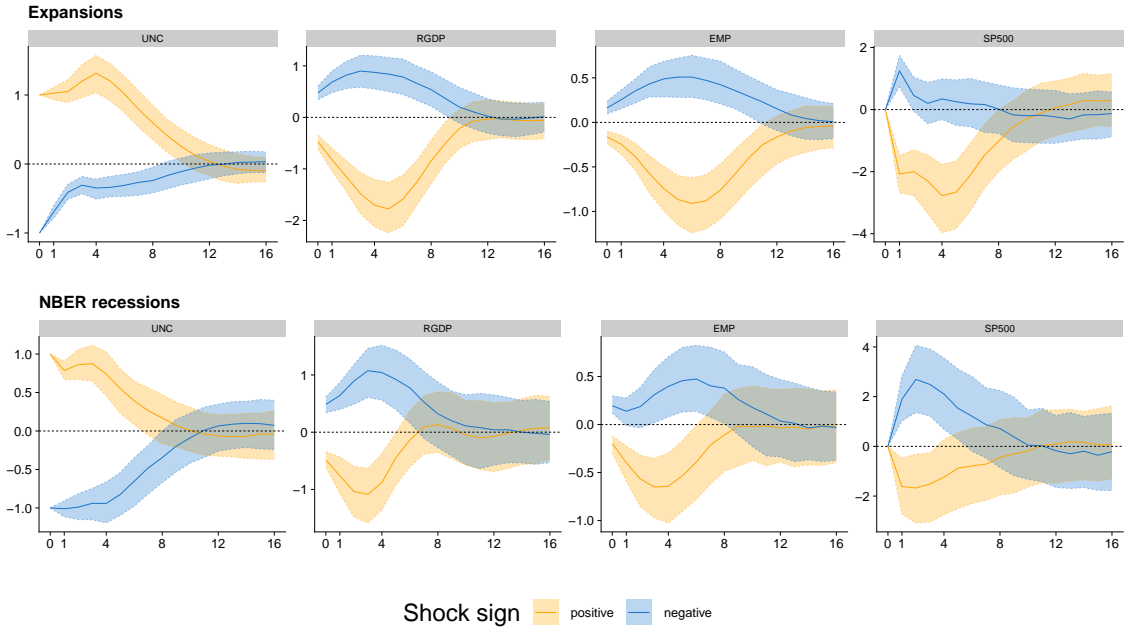
Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. GP-VAR-8 refers to the smallest variant of our non-parametric model and BVAR-8 refers to a small-scale BVAR with SV, which is closely related to the specification used in Jurado, Ludvigson, and Ng (2015).

Figure C.5: Impulse responses of non-focus variables in the GP-VAR-8 for three different sub-sample periods.



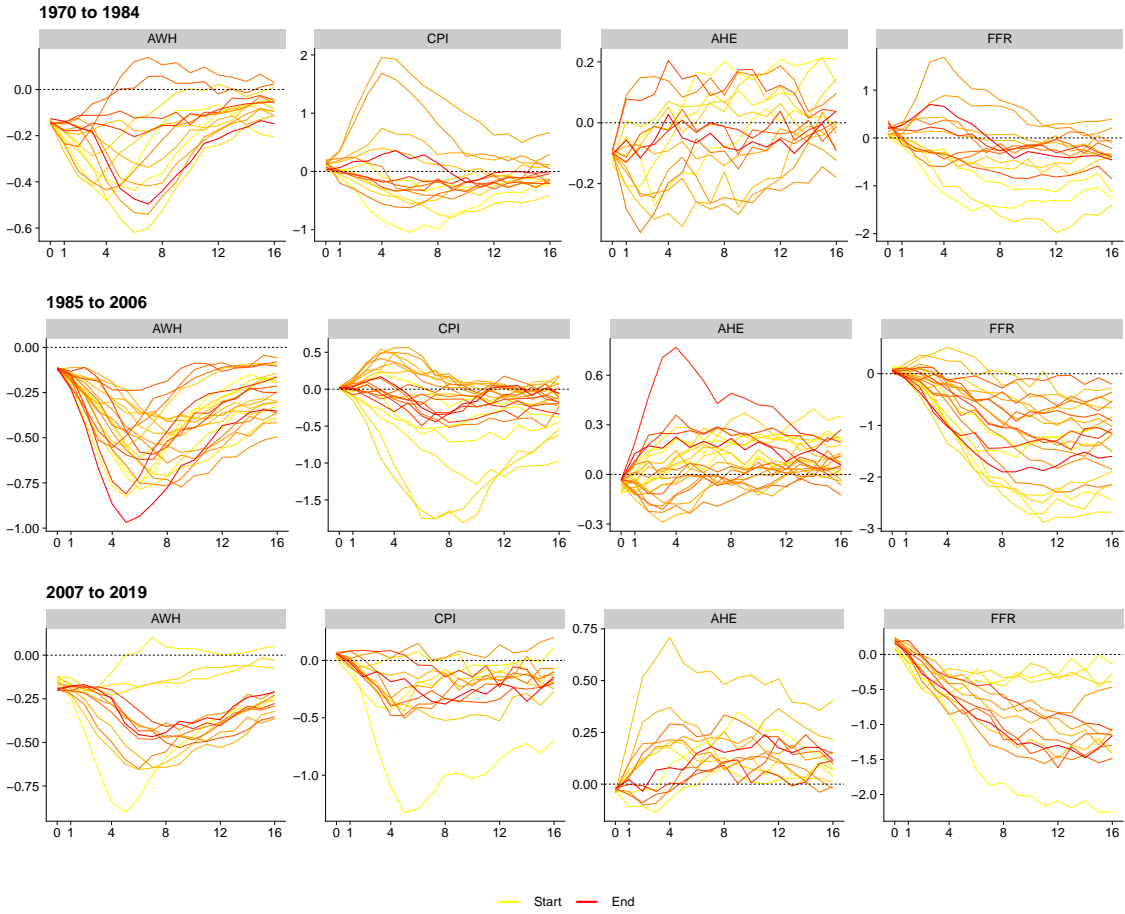
Notes: Period-specific average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive (negative) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets.

Figure C.6: Impulse responses of focus variables in the GP-VAR-8 in recessions and expansions.



Notes: Period-specific generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive (orange) and negative (blue) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. We consider economic recessions and expansions according to the NBER Business Cycle Dating Committee.

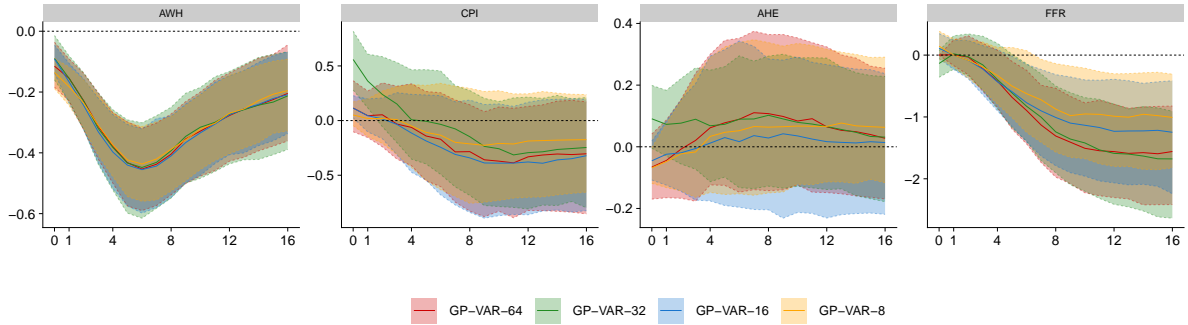
Figure C.7: Period-specific impulse responses of non-focus variables in the GP-VAR-8 across different sub-sample periods.



Notes: Impulse response functions to a positive one standard deviation shock in macroeconomic uncertainty in the GP-VAR-8 across sub-sample periods. Solid lines denote the yearly averaged posterior medians, with colors ranging from yellow (start of the sample) to red (end of the sample).

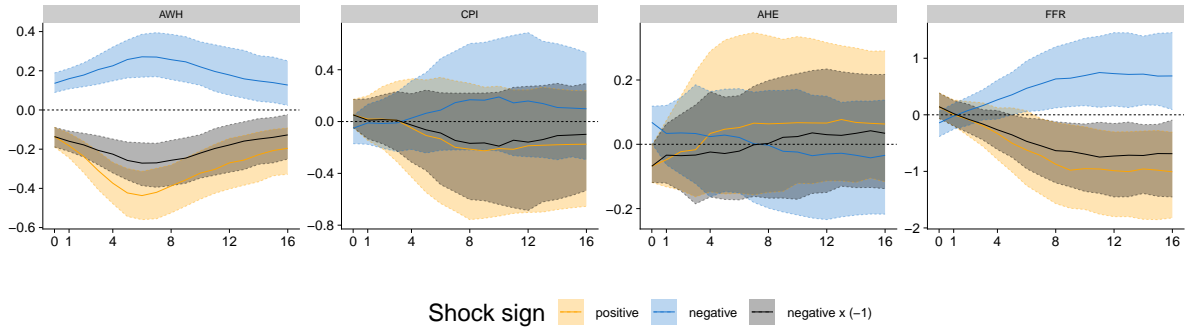
C.4 Additional results on model size and asymmetries

Figure C.8: Impulse responses of non-focus variables across different information sets.



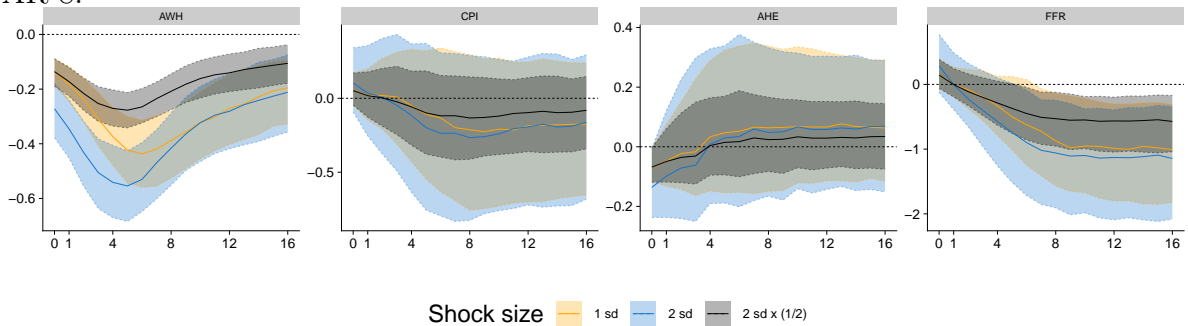
Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets.

Figure C.9: Shock sign asymmetries in responses of non-focus variables for the GP-VAR-8.



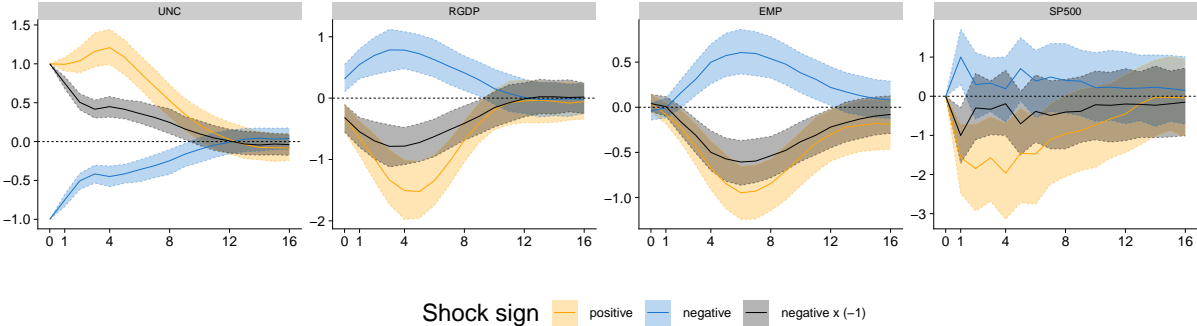
Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a negative (positive) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, negative $\times(-1)$ denotes a negative one standard standard deviation shock with the respective responses being mirrored across the x-axis.

Figure C.10: Shock size asymmetries in responses of non-focus variables for the GP-VAR-8.



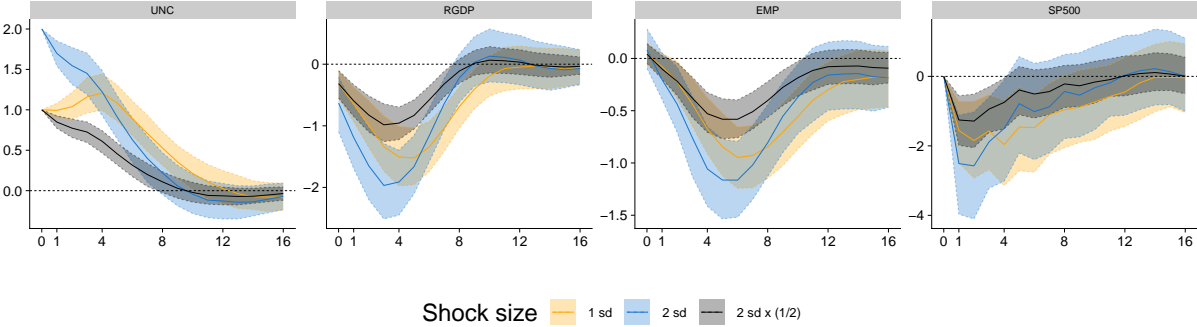
Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive two (one) standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, 1 sd refers to a one standard deviation shock, 2 sd indicates a two standard deviation shock, and $2\text{ sd} \times (1/2)$ denotes a two standard deviation shock with the respective responses divided by two.

Figure C.11: Shock sign asymmetries in impulse responses of focus variables for the GP-VAR-64.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a negative (positive) one standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, negative $x (-1)$ denotes a negative one standard standard deviation shock with the respective responses being mirrored across the x-axis.

Figure C.12: Shock size asymmetries in impulse responses of focus variables for the GP-VAR-64.



Notes: Average generalized impulse responses (GIRFs, outlined in Sub-section 3.5) to a positive two (one) standard deviation shock in macroeconomic uncertainty. Solid lines denote the posterior medians, while shaded areas correspond to the 68% posterior credible sets. Here, 1 sd refers to a one standard deviation shock, 2 sd indicates a two standard deviation shock, and 2 sd $x (1/2)$ denotes a two standard deviation shock with the respective responses divided by two.