Federal Reserve Bank of Cleveland Working Paper Series

A Nonparametric Approach to Augmenting a Bayesian VAR with Nonlinear Factors

Todd E. Clark, Florian Huber, and Gary Koop

Working Paper No. 26-14

June 2026

Suggested citation: Clark, Todd E., Florian Huber, and Gary Koop. 2026. "A Nonparametric Approach to Augmenting a Bayesian VAR with Nonlinear Factors." Working Paper No. 26-14. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202614.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

A Nonparametric Approach to Augmenting a Bayesian VAR with Nonlinear Factors*

Todd E. Clark† Florian Huber‡ Gary Koop§

Abstract

This paper proposes a vector autoregression augmented with nonlinear factors that are modeled nonparametrically using regression trees. There are four main advantages of our model. First, modeling potential nonlinearities nonparametrically lessens the risk of misspecification. Second, the use of factor methods ensures that departures from linearity are modeled parsimoniously. In particular, they exhibit functional pooling where a small number of nonlinear factors are used to model common nonlinearities across variables. Third, Bayesian computation using MCMC is straightforward even in very high-dimensional models, allowing for efficient, equation-by-equation estimation, thus avoiding computational bottlenecks that arise in popular alternatives such as the time-varying parameter VAR. Fourth, existing methods for identifying structural economic shocks in linear factor models can be adapted for the nonlinear case in a straightforward fashion using our model. Exercises involving artificial and macroeconomic data illustrate the properties of our model and its usefulness for forecasting and structural economic analysis.

JEL: C11, C32, C53

KEYWORDS: Nonparametric VAR, nonlinear factor model, regression trees, macroeconomic forecasting, scenario analysis

*We gratefully acknowledge helpful comments from seminar participants at the Central Bank of Colombia. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Please address correspondence to: Florian Huber. Department of Economics, University of Salzburg. *Address*: Mönchsberg 2a, 5020 Salzburg, Austria. *Email*: florian.huber@plus.ac.at. We acknowledge the use of ChatGPT (OpenAI) as an aid for drafting, editing, and brainstorming during the preparation of this manuscript.

†Economist Emeritus, Federal Reserve Bank of Cleveland, and Fellow, Dept. of Economics, Johns Hopkins University

‡Professor of Economics, University of Salzburg

§Professor of Economics, University of Strathclyde

1 Introduction

Researchers in macroeconomics and finance routinely work with high dimensional time series data sets involving dozens or hundreds of variables. Vector autoregressions (VARs), dynamic or static factor models (DFMs or SFMs), and combinations of the two (e.g., factor-augmented VARs or FAVARs) are commonly used. VARs are particularly popular since they often forecast better than factor models (see, among many others, Bańbura, Giannone, and Reichlin, 2010) and a well-developed set of tools exists for carrying out structural economic analysis (e.g., calculating impulse response functions). Bayesian methods are popular in these large, often over-parameterized, models since prior shrinkage can help reduce the problem of over-fitting. Bayesian inference in these models typically proceeds using computationally demanding Markov chain Monte Carlo (MCMC) methods.

Even in the linear homoscedastic VAR or factor model worlds, computational and over-fitting issues can be substantial when the number of variables under study is large (see, e.g., the discussion of VAR computation in Carriero, Clark, and Marcellino, 2019). These problems become greatly magnified in the nonlinear world. Even when the form of nonlinearity is known, incorporating it into a parametric nonlinear VAR can lead to a very non-parsimonious and over-parameterized model. For instance, with VARs, it is common to allow nonlinearities in the form of time-varying parameters (TVP), under which coefficients continuously evolve as random walks. However, carrying out Bayesian inference in a TVP-VAR with as many as 100 variables is a huge computational challenge.[1] Furthermore, in any macroeconomic or financial dataset there are so many empirically plausible nonlinearities that might occur that the researcher who adopts any particular parametric model (e.g., a TVP-VAR or a Markov switching DFM) runs a risk of misspecification.

These considerations motivate the high-dimensional nonlinear time series model we propose in this paper. For reasons detailed below, we view the model and our estimation approach as effectively balancing tradeoffs associated with over-parameterization, misspecification, and computation. One component of the model is simply a linear VAR. Nonlinearities are modeled as deviations from this linear model. In Clark, et al. (2023), this is done for every variable in the system using regression trees. Since this leads to an enormous computational burden and risks over-fitting, we assume that the number of nonlinear functions is much smaller than the number of endogenous series, leading

[1] Chan, Eisenstat, and Strachan (2020) propose using random walk factor structures to reduce the number of processes driving the time-varying parameters and thus speed up estimation.

to favorable computational and statistical properties.

From an econometric perspective, our approach can be seen as featuring functional pooling, with information pooled across variables to improve the estimation of the nonlinear functions determining the factors. Our functional pooling approach extends to VARs with nonlinear functions the pooling of information across series that Müller and Watson (2025) achieve with hierarchical priors in a linear AR model setting.

In light of possible misspecification, inclusion of a nonlinear component (that does not increase the computational burden appreciably) can be interpreted as a defensive modeling strategy (see Linero, 2025), with the nonparametric component serving to control for possible misspecification of the linear model. As discussed in more detail below, our use of regression trees for the nonlinear part of the model is predicated on the previously established empirical success of Bayesian additive regression trees in common macroeconomic datasets.

For the purpose of using the model for structural analysis, we specify the innovation component of the model to include a (linear) static factor model, which permits structural inference through sign restrictions along the lines of Baumeister and Hamilton (2015) and Korobilis (2022).[2] This piece of our specification also contributes to computational efficiency; conditional on the linear static factors (as well as the nonlinear dynamic factors), the VAR block of the model can be estimated equation-by-equation. Together, the factor structures of our model make it scalable to large variable sets.

We consider simulations and a real data analysis. In simulations, our approach produces much better forecasts than linear models when the DGP is nonlinear, while nearly matching the linear VAR benchmark when it is not. We also show that our approach successfully discriminates linear from nonlinear equations. In the real data analysis, we apply our nonparametric nonlinear VAR-factor model to forecast key US macroeconomic aggregates and to analyze the impacts of financial conditions and monetary policy shocks. The forecasting application shows noticeable gains in out-of-sample density forecast accuracy over a linear VAR. The structural application, using sign restrictions for identification, uncovers significant asymmetries in the responses of most indicators of economic activity and inflation: large tightenings of financial conditions and monetary policy

[2]This assumption is consistent with papers such as Bai and Ng (2007); Forni, et al. (2025), who postulate that the economy is driven by a small number of fundamental shocks.

have bigger absolute impacts than easings.

The paper proceeds as follows. Section 2 presents our proposed model, discusses identification and provides additional information on the prior setup. Section 3 provides Monte Carlo evidence on the ability of the model to uncover various types of nonlinearity in common factors. Sections 4 and 5 provide application results on, respectively, out-of-sample forecasting and structural analysis. Section 6 summarizes and concludes. An online appendix provides additional details on data and estimation.

2 Econometric framework

2.1 The factor-BART VAR

Our aim is to model an $M-$dimensional vector $\boldsymbol{y}_t$ that includes macroeconomic and financial variables with a flexible multivariate time series model. Existing nonlinear models (see Primiceri, 2005; Cogley and Sargent, 2005; Sims and Zha, 2006; D'Agostino, Gambetti, and Giannone, 2013; Huber and Rossini, 2022; Huber, et al., 2023, among many others) have the shortcoming that they do not scale well to large dimensions and are prone to over-fitting. There are also recent papers that propose nonparametric Bayesian factor models (see, e.g., Velasco, 2025; Chernis, et al., 2025). These papers assume nonparametric observation equations while restricting the state equations of the factors to be linear. The model we propose differs since the number of unknown functions to estimate is much smaller than M.

Prior regularization in these high-dimensional models is highly effective in lowering the likelihood of over-fitting, but imposing constraints can also be beneficial when they are supported by empirical evidence or theory. This is the route we take in this paper, relying on a nonlinear factor structure. The nonlinearities in $\boldsymbol{y}_t$ may be similar within certain groups of variables. For example, time series capturing economic output may share comparable nonlinear relationships with their lagged fundamentals. Rather than estimating a separate nonlinear function for each variable, it can therefore be more efficient to estimate a single unknown function and assume that, within the class of output measures, the variable-specific functions are all proportional to this common function. This motivates the construction of a model that introduces nonlinear features at the group level of variables, instead of at the level of each variable individually.

The factor-BART VAR assumes that $\boldsymbol{y}_t$ evolves according to the following nonparametric model:

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{A}$ is an $M \times K (= Mp)$ matrix that linearly links $\boldsymbol{y}_t$ to its p lags, stored in $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$.

Nonlinear dynamics are captured by a sequence of Q_μ latent factors contained in the vector $\boldsymbol{\mu}(\boldsymbol{x}_t) = (\mu_1(\boldsymbol{x}_t), \ldots, \mu_{Q_\mu}(\boldsymbol{x}_t))'$. Each component function $\mu_j : \mathbb{R}^K \to \mathbb{R}$ takes $\boldsymbol{x}_t$ as input and returns a scalar output. The $M \times Q_\mu$ coefficient matrix $\boldsymbol{\Lambda}_\mu$ then maps the factors back into the observed series. A key point to note here is that the functional form of $\mu_j(\boldsymbol{x}_t)$ and $\mu_i(\boldsymbol{x}_t)$ can differ. The loadings $\boldsymbol{\Lambda}_\mu$ are then used to weight each of the functions such that the dynamics for a particular variable are best explained. To understand this more clearly, the s^{th} equation of Eq. (1) can be written out as:

$$y_{st} = \boldsymbol{a}'_s \boldsymbol{x}_t + \lambda_{\mu,s1}\mu_1(\boldsymbol{x}_t) + \cdots + \lambda_{\mu,sQ_\mu}\mu_{Q_\mu}(\boldsymbol{x}_t) + \varepsilon_{st}, \tag{2}$$

with $\boldsymbol{a}_s$ denoting the s^{th} row of $\boldsymbol{A}$ and $\lambda_{\mu,sj}$ denoting the $(s,j)^{th}$ element of $\boldsymbol{\Lambda}_\mu$.

Our model breaks the conditional mean for each variable into a parametric part (i.e., a linear regression component) and a nonparametric part (i.e., the nonlinear factors) that is modeled using BART. In macroeconomics, the most popular approach to modeling $\boldsymbol{y}_t$ would be a VAR that sets $\boldsymbol{\Lambda}_\mu = \boldsymbol{0}$. However, this is only optimal if the DGP is linear. If there is evidence that this condition does not hold, we end up with a misspecified model and the estimates of $\boldsymbol{A}$ could lead to misleading inference. As noted above, the presence of the factor can also be interpreted as a defensive modeling strategy (see Linero, 2025), meaning that the nonparametric component serves to control for possible misspecification of the linear model. From this perspective, the main advantage of the factor structure is that similar types of variables might also exhibit similar forms of misspecification of the linear model and our approach can pick this up through the factor loadings.

We assume that the shocks in $\boldsymbol{\varepsilon}_t$ are Gaussian with a covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. Specific assumptions about $\boldsymbol{\Sigma}_\varepsilon$ will be discussed below. Of course, this Gaussian specification constitutes a parametric choice, which (for reasons that will become clear below) is convenient for structural inference with the model. The specification of the model's innovation term could be made less

parametric, such as with an additional BART specification (something considered in some of the models of Clark, et al. (2023)) or infinite mixtures of Gaussians (see, e.g., Huber and Koop, 2024). However, such a specification would make structural analysis more difficult. In any event, with the considerable flexibility afforded by the nonlinear $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ component of the model, we characterize our specification as nonparametric.[3]

The factor-BART VAR specification nests two special cases. First, if $\lambda_{\mu,s1} = \cdots = \lambda_{\mu,sQ_\mu} = 0$, we end up with a linear regression model for equation s. This observation allows us to design a Bayesian shrinkage prior that allows for a data-based assessment of whether nonlinear features are necessary. Second, if we set $\boldsymbol{\Lambda}_\mu = \boldsymbol{I}_M$ (and thus set $Q_\mu = M$), we end up with the mixBART specification proposed in Clark, et al. (2023). Relative to the latter model, our proposed specification is very parsimonious because instead of having to estimate M different functions, we only need to estimate $Q_\mu (\ll M)$ functions.[4] If $M = 100$ and $Q_\mu = 5$, we are estimating only five unknown functions instead of 100, reducing functional complexity by 95% (below we quantify CPU time advantages). This assumption is motivated by the fact that if series display similar forms of nonlinearities (such as different measures of aggregate inflation or subcomponents of GDP), it makes sense to introduce structure on the functional relationship and we achieve this through the factor specification that leads to function pooling.

We illustrate how function pooling works through three simple examples. All three analytical examples assume $M = 3$ observed time series and $Q_\mu = 2$ latent nonlinear factors so that $\boldsymbol{\mu}(\boldsymbol{x}_t) = (\mu_1(\boldsymbol{x}_t), \mu_2(\boldsymbol{x}_t))'$.

Example 1. *Consider the following matrix of factor loadings:*

$$\boldsymbol{\Lambda}_\mu = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \textit{then} \quad \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) = \begin{bmatrix} \mu_1(\boldsymbol{x}_t) \\ \mu_1(\boldsymbol{x}_t) \\ \mu_2(\boldsymbol{x}_t) \end{bmatrix}.$$

[3] Our use of that characterization is consistent with Linero's (2025) characterization of a similar model.

[4] Baumeister, et al. (2025) consider a yet more restricted model in a different setting. In a dynamic panel model, they assume that a single factor determines the nonlinearities that directly impact a set of microeconomic variables, with loadings associated with this factor driven by observed characteristics. The moments of the microeconomic indicators in turn impact the dynamics of macroeconomic variables.

This example implies that series 1 and 2 share the same nonlinear function $\mu_1(\boldsymbol{x}_t)$, while the third series depends on a separate nonlinear function $\mu_2(\boldsymbol{x}_t)$. This setup captures common nonlinear behavior among subsets of series.

Example 2. *Now assume:*

$$\boldsymbol{\Lambda}_\mu = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \textit{then} \quad \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) = \begin{bmatrix} \mu_1(\boldsymbol{x}_t) \\ \mu_2(\boldsymbol{x}_t) \\ 0 \end{bmatrix}.$$

Series 1 and 2 each have their own nonlinear functions, while series 3 exhibits only linear dynamics. The model flexibly allows for nonlinearities to be omitted when not supported by the data.

Example 3. *Finally, suppose:*

$$\boldsymbol{\Lambda}_\mu = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}, \quad \textit{then} \quad \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) = \begin{bmatrix} 0.5\mu_1(\boldsymbol{x}_t) + 0.5\mu_2(\boldsymbol{x}_t) \\ 0.5\mu_1(\boldsymbol{x}_t) - 0.5\mu_2(\boldsymbol{x}_t) \\ \mu_2(\boldsymbol{x}_t) \end{bmatrix}.$$

All three series depend on the same two latent functions, but through different linear combinations. This introduces a complex nonlinear structure while maintaining parsimony if each μ_j takes a simple form.

2.2 Modeling the error variance and approximating the unknown functions

Two components of our model still require discussion. First, the error variance $\boldsymbol{\Sigma}_\varepsilon$. For this, we introduce a factor model so that:

$$\boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Lambda}_q \boldsymbol{\Lambda}_q' + \boldsymbol{\Omega}, \tag{3}$$

with $\boldsymbol{\Lambda}_q$ denoting an $M \times Q_q$ matrix of factor loadings, and $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_M^2)$ being a diagonal matrix of measurement error variances.

Notice that Eq. (3) means the innovation term of the model can be equivalently written as:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Lambda}_q \boldsymbol{q}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \quad \boldsymbol{q}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_q), \tag{4}$$

where $\boldsymbol{q}_t$ are zero mean static factors. This representation shows that the reduced-form errors of the model can be decomposed into a common component $\boldsymbol{\Lambda}_q \boldsymbol{q}_t$ and a sequence of idiosyncratic measurement errors. If we wish to use the model for structural inference, restrictions can be used on $\boldsymbol{\Lambda}_q$ to identify the factors in $\boldsymbol{q}_t$ as fundamental economic shocks as in Korobilis (2022). The assumption that the factors have unit variance fixes their scaling.

Second, we need to discuss how we approximate the unknown functions μ_i. In principle, we can use any nonlinear learning technique such as neural networks (see, e.g. Farrell, Liang, and Misra, 2021; Hauzenberger, et al., 2025a), Gaussian processes (see, e.g., Williams and Rasmussen, 2006; Fox and Dunson, 2015; Tang, Mak, and Dunson, 2024; Hauzenberger, et al., 2025b; Chernis, et al., 2025), or splines (see, e.g. Shin, Bhattacharya, and Johnson, 2020). In this paper, because of its empirical success with many different datasets (including datasets commonly employed in macroeconomics and finance) and the existence of a set of hyperparameters that work well in all of these, we approximate the functions μ_i through Bayesian additive regression trees (Chipman, George, and McCulloch, 2010). BART is a sum-of-trees model that approximates:

$$\mu_i(\boldsymbol{x}_t) \approx \sum_{s=1}^{S} \mathfrak{t}(\boldsymbol{x}_t | \boldsymbol{m}_{i,s}, \mathcal{T}_{i,s}), \tag{5}$$

where $\mathfrak{t}$ is a tree function, $\boldsymbol{m}_{i,s}$ a vector of terminal node parameters, and $\mathcal{T}_{i,s}$ a tree structure and S is the total number of trees. We set $S = 250$ in all applications. These tree structures consist of decision rules that take the form $\{x \leq c\}$ or $\{x > c\}$ and hence split the input space defined by x into a sequence of disjoint sets. For each of these disjoint sets, there is a terminal node parameter that plays the role of a fitted value in a regression model. Trees are prone to over-fitting if no regularization is introduced. Summing over many (possibly) complex trees further increases the risk of over-fitting. As a solution, BART uses Bayesian regularization priors to force each of the trees to take a particularly simple form and thus act as a weak learner. Summation over many such simple trees produces a model with a great deal of representation flexibility, while regularization to keep the trees simple limits the risk of over-fitting.

The factor-BART VAR is capable of capturing a wide range of possible nonlinear relations through the factor structure. As noted in the preceding examples, one possible model formulation could feature common nonlinear behavior among subsets of the data series included in the model,

with most series sharing the same nonlinear function while one or a few series depend on a separate nonlinear function. Alternatively, the nonlinear factor structure might not be full rank, with all series depending on the same small number of latent functions but through different linear combinations. In yet other options (employed in our Monte Carlo simulations presented in the next section), the model might feature (i) nonlinearities arising with threshold effects, so that above and below a certain threshold, the dynamics of the common factors are governed by different sets of coefficients, or (ii) factors that are governed by highly nonlinear functions, such as trigonometric functions. As this discussion suggests, an advantage of using BART or other nonparametric methods is that the range of nonlinear forms it is capable of modeling is enormous. In contrast, Guerrón-Quintana, Khazanov, and Zhong (2023) is an example of a nonlinear dynamic factor model where the nonlinearities involve a particular quadratic form. They show how their model is closely related to a nonlinear DSGE model. Quadratic forms such as this are easily captured using BART.

2.3 Identification

Our model has several latent components that give rise to identification issues. In this section, we discuss how all of them can be solved. The first identification issue arises from the fact that a nonlinear specification nests a linear one and thus it would be possible to obtain $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) = \boldsymbol{A}^* \boldsymbol{x}_t$ so that the nonparametric part of the model might soak up linear effects. In such a case equation (1) becomes:

$$\boldsymbol{y}_t = (\boldsymbol{A} + \boldsymbol{A}^*)\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t.$$

However, the fact that we assume that $Q_\mu < M$ solves this identification problem. This is formalized in Theorem 1.

Theorem 1 (Absorption of $\boldsymbol{A}\boldsymbol{x}_t$)**.** *Let $\boldsymbol{Z} = (\boldsymbol{A}^*\boldsymbol{x}_1, \ldots, \boldsymbol{A}^*\boldsymbol{x}_T)$ be an $M \times T$ matrix with $\boldsymbol{A}^*$ being an $M \times K$ matrix of VAR coefficients with full rank. If $rank(\boldsymbol{Z}) > Q_\mu$, then there is no function $\boldsymbol{\mu}(\boldsymbol{x}_t)$ such that $\boldsymbol{A}^*\boldsymbol{x}_t = \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ for all t. Hence, as long as $rank(\boldsymbol{Z}) > Q_\mu$, the nonlinear factor term $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ cannot absorb the linear VAR component $\boldsymbol{A}^*\boldsymbol{x}_t$.*

Proof. $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ is in the column space of $\boldsymbol{\Lambda}_\mu$, denoted by $\mathcal{C}(\boldsymbol{\Lambda}_\mu)$, which has a dimension of at most Q_μ : dim $\mathcal{C}(\boldsymbol{\Lambda}_\mu) \leq Q_\mu$. If we assume that $\boldsymbol{A}^*\boldsymbol{x}_t = \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ then $\boldsymbol{A}^*\boldsymbol{x}_t \in \mathcal{C}(\boldsymbol{\Lambda}_\mu)$. Hence, $rank(\boldsymbol{Z}) = dim(span\{\boldsymbol{A}^*\boldsymbol{x}_t\}) \leq Q_\mu$. But this is a contradiction since $rank(\boldsymbol{Z}) > Q_\mu$. Hence, no

such $\boldsymbol{\mu}(\boldsymbol{x}_t)$ can exist. □

The condition $rank(\boldsymbol{Z}) > Q_\mu$ depends on the rank of $\boldsymbol{A}^*$ and the rank of the regressor matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)'$. This condition will hold generically if $\boldsymbol{X}$ has sufficient variation and $rank(\boldsymbol{A}^*) > Q_\mu$. The following example illustrates this.

Example 4. *Consider Example 1 and let* $\boldsymbol{a}_j$ $(j = 1, 2, 3)$ *denote the* j^{th} *row of* $\boldsymbol{A}$. *Suppose that* $\boldsymbol{\Lambda}_\mu \mu(\boldsymbol{x}_t)$ *can mimic the linear part of the model. In this case, we have that:*

$$\begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \boldsymbol{a}_3' \end{bmatrix} \boldsymbol{x}_t = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1(\boldsymbol{x}_t) \\ \mu_2(\boldsymbol{x}_t) \end{bmatrix}.$$

A unique solution for $\boldsymbol{A}$ *only exists if* $\boldsymbol{a}_1$ *and* $\boldsymbol{a}_2$ *are proportional to each other, and hence* $rank(\boldsymbol{A}) = 2$.

The non-identified $Q_\mu = M$ case is considered in Linero (2025), who derives theoretical results indicating its attractive properties. For instance, it has excellent contraction properties. And it is worth noting that Linero (2025) shows that these properties hold even if you do not impose orthogonality of the parametric and nonparametric parts of the model. We take these theoretical results as additional justification for this specification choice, even if Q_μ approaches M.

So far we have shown that the model defined above can discriminate between linear and nonlinear effects. However, the factor loadings $\boldsymbol{\Lambda}_\mu$ and factors $\boldsymbol{\mu}(\boldsymbol{x}_t)$ are not separately identified. This is the standard identification issue that also occurs with linear factor models and arises since $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t) = \boldsymbol{\Lambda}_\mu \boldsymbol{R}\boldsymbol{R}' \boldsymbol{\mu}(\boldsymbol{x}_t)$ for any $\boldsymbol{R}$ that satisfies $\boldsymbol{R}\boldsymbol{R}' = \boldsymbol{I}$. This lack of identification gives rise to non-identification of the scale and sign of factors and factor loadings. It also implies that one can permute the columns of $\boldsymbol{\Lambda}_\mu$ and $\boldsymbol{\mu}(\boldsymbol{x}_t)$ without changing the likelihood.

For some purposes this lack of identification poses no problems. That is, identification of this sort is not required for forecasting or impulse response analyses, both of which are functions of the product $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$. Nor is separate identification of factor and factor loadings required to produce partial dependence plots as we do in our empirical work. However, if one is interested in $\boldsymbol{\Lambda}_\mu$ or $\boldsymbol{\mu}(\boldsymbol{x}_t)$ separately, identification restrictions are necessary. These can be implemented in various ways. For instance, similarly to Bernanke, Boivin, and Eliasz (2005), we could restrict the upper $Q_\mu \times Q_\mu$

part of $\mathbf{\Lambda}_\mu$ to an identity matrix. This restriction implies that $\mu_1(\boldsymbol{x}_t)$ is the nonlinear component of series 1, $\mu_2(\boldsymbol{x}_t)$ the nonlinear part of series 2, and so on. For series $j > Q_\mu$ we then again allow for linear combinations of all the functions. This identification scheme has the shortcoming that it leads to order-dependence in terms of how the elements in $\boldsymbol{y}_t$ are ordered. To fix the ordering, one can use economic intuition. For instance, in the case that $Q_\mu = 1$, if one plans to model GDP and its subcomponents, a natural ordering puts GDP on top so that the corresponding value of $\mu_1(\boldsymbol{x}_t)$ would imply a nonlinear relationship between GDP and $\boldsymbol{x}_t$, whereas the subcomponents of GDP have nonlinear relations that are proportional to $\mu_1(\boldsymbol{x}_t)$.

Another approach (which we recommend) is to carry out posterior simulation in the unidentified model (with no restrictions on either $\mathbf{\Lambda}_\mu$ or $\boldsymbol{\mu}(\boldsymbol{x}_t)$) and then identify the sign and scale of loadings and nonlinear factors a posteriori while using a non-exchangeable prior on the factor loadings to fix column switching (see Section 2.4). Fixing the scale can be achieved by computing the standard deviation of $\boldsymbol{\mu}_j = (\mu_j(\boldsymbol{x}_1), \ldots, \mu_j(\boldsymbol{x}_T))'$, labeled $\sigma_{\mu,j}$, for all j and then computing (for each MCMC draw):

$$\mathbf{\Lambda}_\mu^* = \mathbf{\Lambda}_\mu \text{ diag}(\sigma_{\mu,1}, \ldots, \sigma_{\mu,Q_\mu}) \quad \text{and} \quad \boldsymbol{M}^* = \boldsymbol{M} \text{ diag}(\sigma_{\mu,1}^{-1}, \ldots, \sigma_{\mu,Q_\mu}^{-1}),$$

where $\boldsymbol{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{Q_\mu})$ is a $T \times Q_\mu$ matrix. This implies that the columns of $\boldsymbol{M}^*$ have variance 1 and thus the variance of the nonlinear factors is absorbed by the factor loadings and it preserves the product $\mathbf{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$ and hence leaves the likelihood unchanged.

Given $\mathbf{\Lambda}_\mu^*$ and $\boldsymbol{M}^*$, we follow Kastner, Frühwirth-Schnatter, and Lopes (2017) and Kastner (2019) and search for the series that has a posterior distribution of the loadings that is bounded away from zero. Then, for each $\mu_j(\boldsymbol{x}_t)$ we search through each series and identify the series whose smallest absolute posterior draw is largest and fix the corresponding sign to be positive. This rule selects a pivot series whose loading has posterior support away from zero, thereby anchoring the sign of each factor across MCMC draws. Formally, if $s = 1, \ldots, S$ are the post burn-in draws from the posterior, we assign a positive sign to $|\lambda_{ij,\mu}^*|$ as:

$$\underset{i \in \{1,\ldots,M\}}{\arg\max} \left(\min_{s \in \{1,\ldots,S\}} |\lambda_{ij,\mu}^*| \right).$$

The final identification issues are related to $\mathbf{\Lambda}_q$ and $\boldsymbol{q}_t$ and its interaction with $\mathbf{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$.

These are sorted out by fixing the variance of $\boldsymbol{q}_t$ to $\boldsymbol{I}_{Q_q}$, using sign restrictions on the loadings (see sub-section 5.2) and noting that $\boldsymbol{\mu}(\boldsymbol{x}_t)$ is a deterministic function of lagged observables while $\boldsymbol{q}_t$ is a sequence of white noise shocks. Hence, nonlinear factors cannot absorb the shock factors (and vice versa).

2.4 Priors

Our model involves different blocks of parameters or latent states (e.g., the VAR block, the factor loadings, and the BART factors). For each of these blocks, earlier work in the literature has proposed a range of different priors and any combination of these can be used with our factor-BART VAR. We make particular choices from this set of standard priors. On the VAR parameters we use a horseshoe prior (Carvalho, Polson, and Scott, 2009) with row-specific global shrinkage terms; on the BART factors we use precisely the same prior setup proposed in Chipman, George, and McCulloch (2010); on the ω_j^2 we use inverse Gamma priors; and on $\boldsymbol{\Lambda}_q$ we use either truncated Gaussian priors (as discussed in Section 5) or weakly informative Gaussian priors.

The only departure from standard choices is the prior we use on $\boldsymbol{\Lambda}_\mu$. In this case, we consider a row- and column-wise shrinkage prior (see, e.g., Huber and Feldkircher, 2019; Kastner, 2019). Particularly, we consider a shrinkage prior that has row- and column-specific shrinkage terms:

$$\lambda_{\mu,ij} \sim \mathcal{N}(0, \psi_{\mu,ij}^2 \tau_{\mu,i}^2 \varpi_j), \quad \psi_{\mu,ij} \sim \mathcal{C}^+(0,1), \quad \tau_{\mu,i} \sim \mathcal{C}^+(0,1),$$
$$\varpi_j = \frac{\varpi}{j^2}, \quad \varphi \sim \mathcal{G}^{-1}(a_\varpi, b_\varpi),$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, Q_\mu$ where $\mathcal{C}^+(0,1)$ denotes the half-Cauchy distribution and $\mathcal{G}^{-1}(a_\varpi, b_\varpi)$ the inverse Gamma distribution with hyperparameters a_ϖ, b_ϖ. We set these equal to $a_\varpi = 3$ and $b_\varpi = 0.03$. Conditional on ϖ_j, this is a standard horseshoe prior with row-specific global shrinkage terms. The row-wise scaling parameters $\tau_{\mu,i}$ shrink a particular row toward zero. Setting $\tau_{\mu,i} \approx 0$ would therefore imply that the resulting i^{th} equation is linear. The presence of the local shrinkage parameters $\psi_{\mu,ij}^2$ effectively allows for selecting a particular nonlinear form, if warranted by the data.

This prior also serves as a device for selecting the effective number of nonlinear factors by increasingly forcing the columns in $\boldsymbol{\Lambda}_\mu$ to zero. The presence of the local scaling terms then allows

for deviations from this shrinkage pattern, if necessary. This approach automatically removes unnecessary factors from the model by shrinking factor loadings associated with irrelevant factors to zero.

Our model can be estimated using a sophisticated Markov chain Monte Carlo algorithm that involves sampling from (mostly) well-known full conditional posterior distributions. The detailed algorithm and all full conditional distributions are provided in the online appendix.

2.5 Computational efficiency

To flesh out the discussion above of the computational efficiency that comes with the factor specification of the factor-BART model, we have run some simulations to quantify computational runtimes for models differing in the number of variables M and of common nonlinear factors Q_μ. Figure 1 reports the number of seconds necessary to generate $1,000$ draws from our MCMC algorithm for a dataset (generated randomly) with $T = 500$ observations and for different values of $M \in \{20, 30, 40\}$. The lines give the log times, with the left side of the vertical axis providing times and the right side of the axis indicating how computation times multiply relative to the baseline, the smallest model having $M = 20$ and $Q_\mu = 2$. The slope in the legend refers to the percentage increases in runtimes from including additional factors.

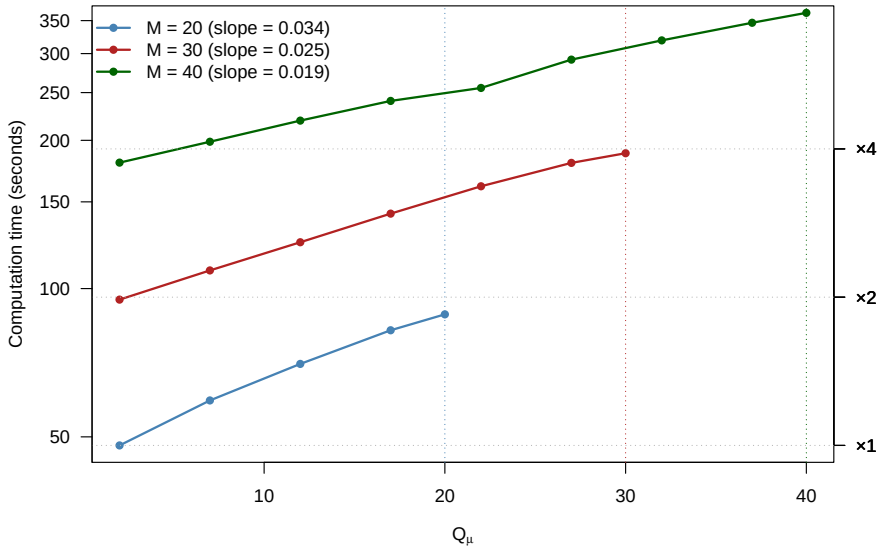


Figure 1: Runtimes to produce 1,000 draws across factor-BART VAR sizes

For a given M, the CPU time to estimate the model rises sharply with the number of BART

terms. For example, with $M = 30$, the CPU time roughly doubles as the number of factors increases from $Q_\mu = 2$ to $Q_\mu = 30$ (i.e., as the model moves from our proposed specification to the mixBART model). However, as the number of model variables (M) rises, for a given set of variables the rate of increase in CPU time with additional factors diminishes somewhat (i.e., the reported slopes of the lines decline with M). This pattern is driven by the fixed costs of the estimation algorithm. Returning to our statement above about scaling advantages that come with the factor-BART specification, the results in Figure 1 indicate that a model with 30 variables and 2 factors can be estimated in little more than the time required to estimate a model with 20 variables and 20 factors.

3 Artificial data exercises

3.1 Design of the Monte Carlo study

In this section, we carry out an empirical exercise involving artificial data. We generate artificial datasets of $M = 16$ time series of length $T \in \{250, 500\}$. The choice of M is moderately large and typical of what is used in the VAR literature. The choices for T are the sort that often occur with quarterly and monthly macroeconomic datasets.[5]

Our first four DGPs use the following general specification:

$$\boldsymbol{y}_t = \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \boldsymbol{\Lambda}_1\boldsymbol{\mu}_t + \boldsymbol{\Lambda}_2\boldsymbol{q}_t + \boldsymbol{\eta}_t,$$

with $\boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, 0.01^2 \times \boldsymbol{I}_{16}), \boldsymbol{q}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_2), \boldsymbol{y}_0 = \boldsymbol{y}_{-1} = \boldsymbol{0}$, and the elements of $\boldsymbol{\Lambda}_1$ being drawn from $\mathcal{N}(0.5, 0.25^2)$ and then sparsified by randomly setting to zero 50% of its rows. Note that this specification implies that half of the equations in our DGPs are linear and do not contain the nonlinear factor. The coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are sparse. Their non-zero elements are drawn from $\mathcal{N}(0, 0.1^2)$ and $\mathcal{N}(0, 0.05^2)$, respectively, with 90% of the entries set to zero at random. The diagonal elements are fixed at $\text{diag}(\boldsymbol{A}_1) = 0.25$ and $\text{diag}(\boldsymbol{A}_2) = 0.10$ to ensure stability. The

[5] In results omitted in the interest of brevity, we also conducted experiments with $T = 10,000$, a sample large enough to effectively make priors irrelevant and ensure that the data determine the posteriors. Qualitatively, results on the MSEs of coefficients for the very large sample are very similar to those reported for $T = 500$, with our model yielding coefficient accuracy gains with nonlinear DGPs and largely matching the accuracy of the linear VAR.

elements of $\mathbf{\Lambda}_2$ are drawn from standard Gaussian distributions, with two diagonal elements set to 1 for scale normalization. We assume $Q_q = 2$. Our fifth DGP departs from the general specification by making $\boldsymbol{\eta}_t$ t-distributed with 4 degrees of freedom, rather than Gaussian.

With this general setup, we consider five specific DGPs. `DGP 1` and `DGP 5` set $\mathbf{\Lambda}_1 = 0$ and, thus, are linear. Accordingly, there is no need to specify the factors. For the remaining DGPs, we assume $Q_\mu = 3$ factors and introduce different laws of motion for their conditional means. They all assume that the factor evolves according to the nonlinear functions of lagged observables.

In particular, the nonlinear factor depends on $\tilde{\boldsymbol{y}}_{t-1} = \frac{1}{2}(\boldsymbol{y}_{t-1} + \boldsymbol{y}_{t-2})$ in three different ways. `DGP 2` sets $\mu(\boldsymbol{x}_t) = \boldsymbol{B}_1 \cos(\tilde{\boldsymbol{y}}_{t-1} \cdot \tilde{y}_{t-1,1} \cdot \tilde{y}_{t-1,3} \cdot \pi) + \boldsymbol{B}_2 |\tilde{\boldsymbol{y}}_{t-1} - 0.5| + \boldsymbol{B}_3 \sin(\tilde{\boldsymbol{y}}_{t-1})$, with operations inside trigonometric functions and the absolute value being element-wise. This DGP is inspired by the statistics literature (see, e.g., Chipman, George, and McCulloch, 2010). In terms of the properties of the simulated time series this DGP implies cyclical movements of the factors (consistent with business cycle fluctuations), with the length and intensity of the waves being driven by lagged endogenous series.

`DGP 3` is an endogenous regime-switching specification, with the switching process depending on the sign of the first component of the smoothed lag vector. Specifically, this DGP sets $\mu(\boldsymbol{x}_t) = [\boldsymbol{B}_1 \tilde{\boldsymbol{y}}_{t-1}] \times \mathbb{I}(\tilde{y}_{1,t-1} < 0) + [\boldsymbol{B}_2 \tilde{\boldsymbol{y}}_{t-1}] \times \mathbb{I}(\tilde{y}_{1,t-1} \geq 0)$. Similar processes have been considered in, e.g., Kolesár and Plagborg-Møller (2025).

`DGP 4` resembles the one studied in, e.g., Gonçalves, et al. (2024) and Hauzenberger, et al. (2025a) and implies size and sign asymmetries in how $\boldsymbol{y}_t$ reacts to lagged values. This DGP sets $\mu(\boldsymbol{x}_t) = \frac{1}{4}\boldsymbol{B}_1 \tilde{\boldsymbol{y}}_{t-1}^3 + \boldsymbol{B}_2 \boldsymbol{r}_t$, where $\tilde{\boldsymbol{y}}_{t-1}^3 = \tilde{\boldsymbol{y}}_{t-1} \odot \tilde{\boldsymbol{y}}_{t-1} \odot \tilde{\boldsymbol{y}}_{t-1}$ (multiplication is applied element-wise) and the j^{th} element of $\boldsymbol{r}_t$ is given by $r_{jt} = \max(0, \tilde{y}_{j,t-1})$ for $j = 1, \dots, 16$.

In all of the nonlinear DGPs, the elements of the matrices $\boldsymbol{B}_j$ in the factor equations are sampled from Gaussian distributions with mean zero and variances for $\boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3$ equal to $0.15^2, 0.075^2$ and 0.1^2, respectively.

In what follows, with each artificial dataset we estimate two models: the factor-BART VAR and a linear VAR that serves as a benchmark. For the factor-BART VAR we set $Q_\mu = 8$ nonlinear factors, which is a much larger value than that used in the DGPs. Thus our estimating model is over-parameterized. We do this so as to investigate the ability of our prior to shrink and choose the correct more parsimonious specification. In both of the models we set $Q_q = 3$, which is larger

than the value used in the DGPs. The linear VAR is identical to the factor-BART VAR except that $Q_\mu = 0$ (in lieu of adding terminology to reflect the linear factor structure that we retain in the model's innovation term, we simply refer to this as the linear VAR). Hence, accuracy differences solely reflect performance gains from adding nonlinear factors.

3.2 Monte Carlo findings

We begin the results of the simulation exercise with out-of-sample forecast accuracy, specifically, the joint log predictive likelihoods (LPLs) reported in Table 1. The table shows the different DGP configurations in the columns. The rows include the joint LPLs across forecast horizons and for different values of T. All numbers are relative to the linear VAR benchmark and are averaged across 500 replications from the DGP.

		DGP				
		1	2	3	4	5
$T = 250$	$h = 1$	0.16	4.94	5.19	6.13	0.02
	$h = 4$	-0.01	3.75	2.43	3.08	0.13
	$h = 8$	0.07	3.40	2.39	4.46	-0.03
$T = 500$	$h = 1$	-0.03	4.46	4.27	6.68	-0.24
	$h = 4$	0.08	3.89	2.89	2.81	-0.14
	$h = 8$	-0.18	3.87	2.35	3.94	-0.33

Notes: The table shows joint log score differences across 500 replications from each of the DGPs to the linear VAR (which sets $Q_\mu = 0$). Numbers greater than zero imply a better performance in terms of LPLs.

Table 1: Forecast performance based on synthetic data for 500 replications from the DGP.

Our main finding is that, for the nonlinear DGPs, our factor-BART VAR forecasts better than the linear benchmark. This is unsurprising, but also illustrates how BART methods can uncover a diverse range of nonlinear forms, including oscillatory, threshold, and asymmetric dynamics. Since BART is a method for classifying observations into groups (as opposed to methods such as Gaussian processes that smooth across local observations), we had expected it to work relatively better for DGPs involving abrupt breaks (such as `DGP 3`) and relatively worse in smooth DGPs (such as `DGP 2`). We do not find this to be the case, with BART performing roughly as well for our three very different nonlinear DGPs. It is also reassuring that even for the linear DGP, `DGP 1`, we find that our factor-BART VAR performs roughly the same (albeit not quite as well, as expected) as the true linear model. Thus, we find that our nonparametric model can approximate linearity without over-

fitting. For the DGP with fat tailed errors, `DGP 5`, both of the estimating models are misspecified with respect to density forecasting. There is very little difference in forecast performance between the two models. But with $T = 500$, the linear VAR forecasts slightly better than our factor-BART VAR. The latter model does not do too well at picking up the fat tails in the errors in the DGP that are missing in both the estimating and the forecasting models.

Next, we consider how well our specification shrinks various equations toward linearity. To this end, we compute the posterior distribution of the area under the receiver operating characteristic curve (AUC) based on the shrinkage parameters of the prior on $\mathbf{\Lambda}_\mu$. To this end, we define a shrinkage score S_j that is given by:

$$S_j = \frac{\sum_{i=1}^{Q_\mu} \log \phi_{ij}^{-1}}{Q_\mu},$$

where $\phi_{ij} = \varpi_{\mu,j}\psi^2_{\mu,ij}\tau^2_{\mu,ij}$ denotes the actual prior variance associated with the $(i,j)^{th}$ element in $\mathbf{\Lambda}_\mu$. Notice that larger values of S_j imply stronger shrinkage toward linearity of the j^{th} equation. Define $\mathcal{Z}$ to be the set of indices for zero rows in $\mathbf{\Lambda}_\mu$ and $\mathcal{S}$ the set of non-zero rows in $\mathbf{\Lambda}_\mu$. Given S_j, the AUC is defined as:

$$\text{AUC} = \frac{1}{|\mathcal{Z}||\mathcal{S}|}\sum_{i\in\mathcal{Z}}\sum_{s\mathcal{S}}\mathbb{I}(S_i > S_s),$$

which approximates the probability that a randomly selected zero row is more strongly shrunk than a randomly selected non-zero row. The AUC is defined as the posterior probability that an equation that is truly linear ($\boldsymbol{\lambda}_j = \mathbf{0}$) receives stronger shrinkage than an equation that is truly nonlinear ($\boldsymbol{\lambda}_j \neq \mathbf{0}$), and therefore measures the ordering (separation) ability of the shrinkage mechanism without requiring a threshold.

We compute the AUC value for every MCMC draw, compute the posterior median of the AUC, and then report boxplots that show the 25th, 50th, and 75th percentiles of posterior medians for the 500 replications from the DGP in Figure 2. Notice that the linear DGPs, `DGP 1` and `DGP 5`, are absent from the figure since the set $\mathcal{Z}$ is empty and the AUC is not defined in this case. This figure clearly shows that the factor-BART VAR classifies the linear equations well and shrinks strongly toward linearity. This holds for all the different types of nonlinearity in our DGPs. For example, with $T = 250$, the AUC is about 0.95 for DGP 2 and 0.85 for DGPs 3 and 4; these are very high posterior probabilities that estimates of linear equations are shrunk more strongly than

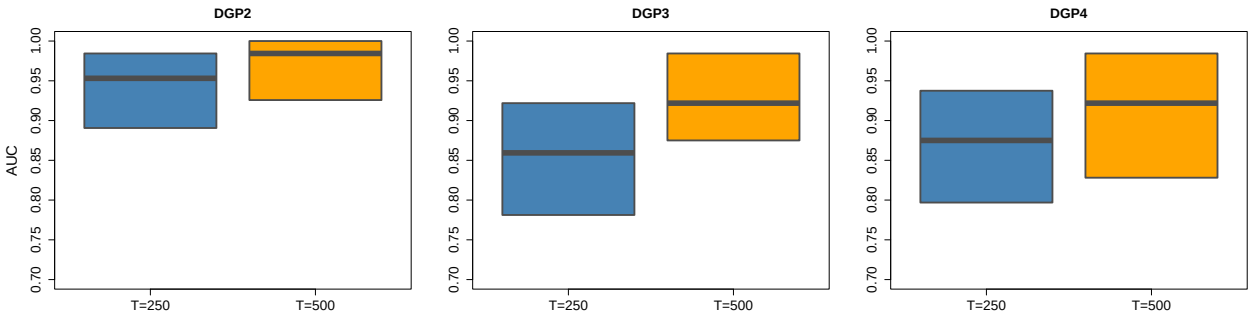


Figure 2: Posterior percentiles of the AUC score

estimates of truly nonlinear equations. Unsurprisingly, larger sample sizes lead to modestly higher AUCs and implied probabilities.

In many applications, the key parameters of interest in our model are the linear VAR coefficients $\boldsymbol{A}$. Estimating a linear VAR if the DGP is nonlinear leads to misspecification and we expect that the accuracy of the posterior estimates of $\boldsymbol{A}$ from the linear model will deteriorate in this case. By contrast, estimating a nonlinear model if the DGP is linear might also have negative effects on the estimates of $\boldsymbol{A}$ due to over-fitting. Given that the AUC results indicate that our model successfully discriminates linear from nonlinear equations, we now ask whether this also leads to improved parameter estimates. To analyze this, we compare (across artificial datasets) the mean squared error (MSE) between the true values of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ and the posterior median of $\boldsymbol{A}$ of the factor-BART VAR to the linear VAR. Table 2 reports the mean MSEs for the factor-BART VAR relative to the linear VAR across 500 replications from each DGP.

	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5
$T = 250$	0.99	0.98	0.94	0.89	0.98
$T = 500$	1.00	0.97	0.94	0.86	0.99

Notes: The table shows MSE ratios across 500 replications from each of the DGPs from factor-BART VAR estimates relative to the linear VAR (which sets $Q_\mu = 0$). Numbers below one imply that the factor-BART VAR provides more accurate estimates than the VAR.

Table 2: MSE between the posterior median of $\boldsymbol{A}$ and the true value of $\boldsymbol{A}$ across 500 replications from the DGP

This confirms that for the nonlinear DGPs, the factor-BART VAR is doing a better job of estimating the coefficients of the linear part of the model. In the case of the nonlinear DGPs 2, 3, and 4, the MSE ratios for the estimates of the coefficients of $\boldsymbol{A}$ are consistently below 1, with improvements ranging from 2% to 14%. Even for the linear DGP, the factor-BART VAR does

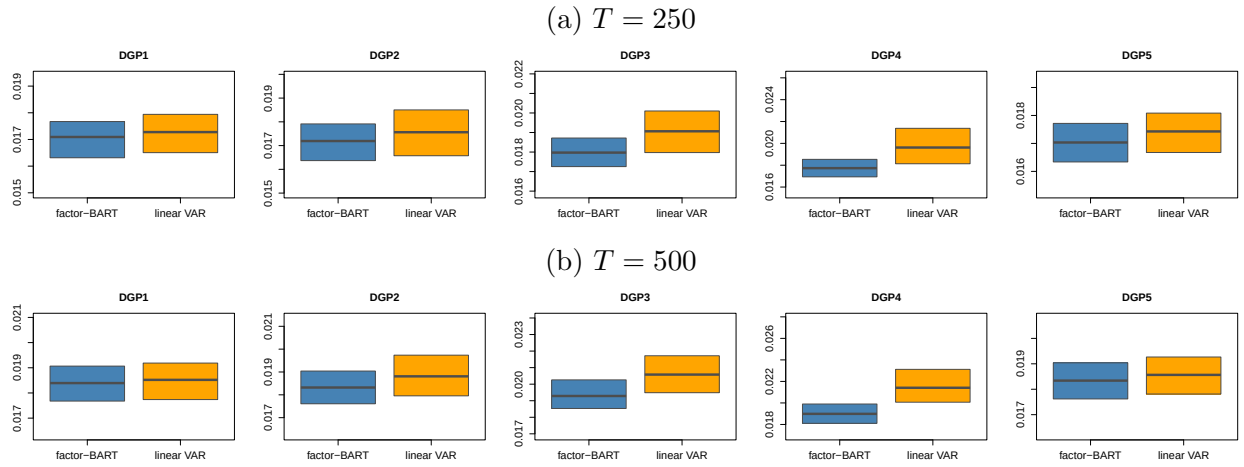


Figure 3: Posterior percentiles of the mean squared errors between the true VAR coefficients and the posterior median of $\boldsymbol{A}$.

roughly as well estimating the linear part of the model as the (true) linear VAR, with MSE ratios very near or at unity. The comparability of the linear VAR and factor-BART VAR estimates with a linear DGP may be connected to priors in estimation and BART. When the DGP is linear, the use of Bayesian shrinkage in estimation induces some bias in the linear VAR's coefficient estimates (specifically in the case of the coefficient on the own first lag in each equation, for which the prior means of 0 differ from the DGP coefficients). In that context, the accuracy of the factor-BART VAR's estimates of the VAR coefficients may be helped by the inclusion of BART terms in the model. Together, on net these forces may contribute to the comparability of coefficient MSEs in DGPs 1 and 5.

To get a sense of the uncertainty surrounding the values in Table 2 that average across all replications, we provide boxplots of the MSEs in Figure 3. In these plots, the black lines and boxes display the 25th, 50th, and 75th percentiles of the factor-BART VAR (blue boxes) and linear VAR (yellow boxes) squared coefficient errors across the 500 simulations. It can be seen that our finding that the factor-BART VAR does a better job of estimating the linear part of the model when the true DGP is nonlinear and does nearly as well when the true DGP is linear holds up robustly across draws from the DGPs. Having demonstrated that our approach performs well in simulations, we now apply it to real-world data.

4 Application: Forecasting the US economy

4.1 Data, design of the forecasting exercise and model specification

In our applications, we use $M = 22$ major quarterly macroeconomic and financial time series. With two exceptions, we take the data from the FRED-QD database developed in McCracken and Ng (2021) and maintained at the Federal Reserve Bank of St. Louis. As detailed in the variable list provided in Table A.1 of the online appendix, the selected series include growth of real GDP and many of its main components, various measures of aggregate inflation, other indicators of the labor market and economic activity such as payroll employment and industrial production, the federal funds rate and various other interest rates covering longer maturities. One exception to the data source is that we also include in the model the excess bond premium (EBP) of Gilchrist and Zakrajšek (2012) (using the current updated series available from Favara, et al., 2016). In the second exception, we include in the model the utilization-adjusted measure of growth in total factor productivity developed in Fernald (2014) (and downloaded from the author's website). Our set of selected variables is very similar to that used in the large Bayesian VAR of Crump, et al. (2025). We transform all variables to be stationary following recommendations from McCracken and Ng (2021); see appendix Table A.1 for details. With transformations implemented, our data sample runs from 1976:Q3 to 2023:Q4.

The forecast design is as follows. We start with an initial estimation sample that ends in 2001:Q4. After estimating the model we iteratively compute predictive distributions from $h = 1$ up to $h = 8$ quarters ahead. We then add one more observation to the model estimation sample (i.e., use a recursive scheme) and repeat this procedure until we reach the end of the sample. Finally, after obtaining the predictive distributions for all observations in the holdout, we compute energy scores to evaluate density accuracy for the full set of variables and continuous ranked probability scores (CRPSs) to evaluate density accuracy on a variable-by-variable basis.[6]

We consider models that set $p = 2$ and $Q_q = 3$ and focus on two values of Q_μ. One sets

[6]The energy score is a generalization of the CRPS, to which it collapses in the univariate case: $\text{ES}_t(\boldsymbol{y}_t) = E_{\mathfrak{f}}||\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t|| - 0.5E_{\mathfrak{f}}||\hat{\boldsymbol{y}}_t - \hat{\boldsymbol{y}}_t'||$, where $\hat{\boldsymbol{y}}_t$ and $\hat{\boldsymbol{y}}_t'$ are independent random vectors with distribution $\mathfrak{f}$. The energy score and the CRPS are less sensitive to outliers (like the extremes experienced during the COVID pandemic) than the LPL.

$Q_\mu = 8$ and is thus relatively large. In this case, we let our shrinkage prior detect the effective number of nonlinear factors. Another model sets $Q_\mu = 3$. This model serves as a benchmark on how much predictive accuracy we lose if we focus on models that are parsimonious in terms of actual nonlinear functions to estimate. As additional competing models, we include the mixBART specification of Clark, et al. (2023), a Gaussian process (GP) VAR (Hauzenberger, et al., 2025b), as well as a time-varying parameter VAR with a horseshoe shrinkage prior on the time-invariant parameters and the state innovation variances.[7] As a benchmark, we consider a linear Bayesian VAR that sets $Q_\mu = 0$ (while keeping $Q_q = 3$, i.e., keeping a linear factor structure in the model's innovation term).

4.2 Predictive evidence

Table 3 provides average energy scores for our factor-BART VAR specification and the alternatives described above, all relative to average scores for the linear VAR. The factor-BART VAR model consistently improves on the accuracy of the linear model, by small margins (1 to 2%) at the one-step horizon but more sizable margins (10 to 13%) at multi-step horizons. Results are largely the same with 3 nonlinear factors as with 8 nonlinear factors. The mixBART specification yields accuracy similar to that of the factor-BART VAR. From the perspective of overall density forecast accuracy, it follows that our introduction of a factor structure to capture nonlinearities rather than allowing a nonlinear component for each variable provides computational gains facilitating large models without reducing forecast accuracy. Of the other specifications considered, the nonparametric alternative of the GP-VAR slightly improves on the linear VAR's accuracy for multi-step horizons but falls short of the accuracy of the BART-based specifications. Finally, the TVP-VAR falls short of the accuracy of all the other models. This may stem from the rich parameterization that comes with using TVPs in a large model.

Of course, while the factor-BART VAR model improves forecast accuracy on average over the 2002-2023 period, performance of the model and the alternatives may vary over the sample. To assess how accuracy changes over the hold-out sample, Figure 4 reports recursive average score ratios (for each model relative to the linear VAR) over time.[8] These results show that the specifications

[7] Our GP VAR model differs from the original implementation of Hauzenberger, et al. (2025b) by featuring a factor structure in the shocks.

[8] At a given date t in the chart, a given line shows the ratio of the energy score for that model from the beginning

	fBART				
$h \downarrow$	$Q_\mu = 8$	$Q_\mu = 3$	mixBART	GP-VAR	TVP-VAR
1	0.98	0.99	1.02	1.02	1.11
4	0.89	0.90	0.90	0.94	1.15
8	0.89	0.90	0.90	0.95	1.36
12	0.87	0.88	0.89	0.94	0.99

Table 3: Energy scores relative to the linear BVAR over the hold-out sample: 2002:Q1 to 2023:Q4

that include a nonlinear component capturing nonlinearities achieve forecast accuracy gains over the linear VAR over time, and not just on average. At the one-step horizon, the (recursive average) gains are more than 10% from 2002 through about 2015, and then decline in the remainder of the sample. At longer horizons, the gains to including BART in the model tend to increase from the Great Recession up until the pandemic. Among the BART specifications, over good parts of the sample, our factor-BART VAR with $Q_u = 8$ factors is slightly more accurate than the same model with 3 factors or mixBART, but as indicated above, on average performance is similar. The alternative GP-VAR and TVP-VAR specifications are clearly less accurate for most of the sample.

To provide a sense of forecast performance across variables, Figure 5 reports relative CRPS results for a selection of key variables, including growth in GDP and industrial production, the unemployment rate, the federal funds rate, and headline CPI and core PCE inflation. For each variable-horizon combination considered, the figure provides the median (black bar) and interquartile range (box) of the time series of the CRPS for each nonlinear model relative to the CRPS of the linear VAR baseline. Broadly, these results confirm the patterns seen in the energy scores for the full set of model variables. Except for the inflation variables, the specifications including a nonlinear BART component improve on the accuracy of the linear VAR (and in most cases do so across all horizons). Our factor-BART VAR once again achieves accuracy gains similar to those achieved with the mixBART model, but with reduced computation requirements. For these variables, forecasts from the GP-VAR and TVP-VAR models are modestly to substantially less accurate than forecasts from the BART models and the linear VAR. In the case of the inflation measures, the factor-BART VARs yield forecasts more accurate than those from all the other nonlinear specifications, including mixBART. However, for these variables, the factor-BART VAR is not quite as accurate as a linear VAR.

of the sample through t relative to the score for the linear VAR over the same period.

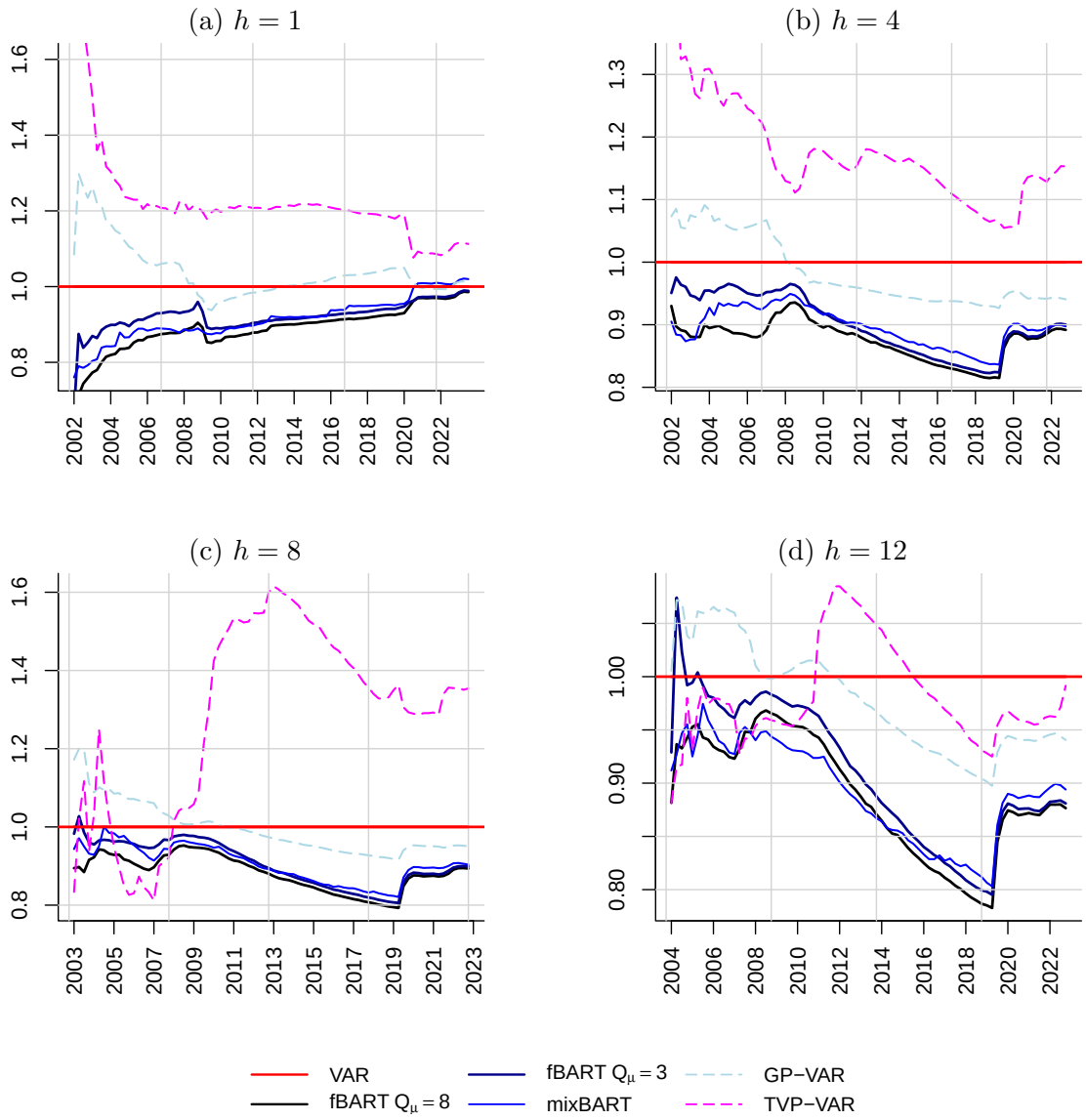


Figure 4: Evolution of the average energy score over the hold-out period.

5 Structural analysis using a large nonlinear model of the US economy

Having established some benefits of our proposed model for out-of-sample forecasting, this section's application shows its ability to capture nonlinearities in responses to structural shocks. In this application, we use precisely the same set of $M = 22$ variables and factor-BART VAR specification as in the forecasting application (with $Q_\mu = 8$). We use sign restrictions (detailed below) as in studies such as Korobilis (2022) to identify shocks to demand, supply, monetary policy, and financial conditions. Accordingly, we set the number of common shocks to $Q_q = 4$.

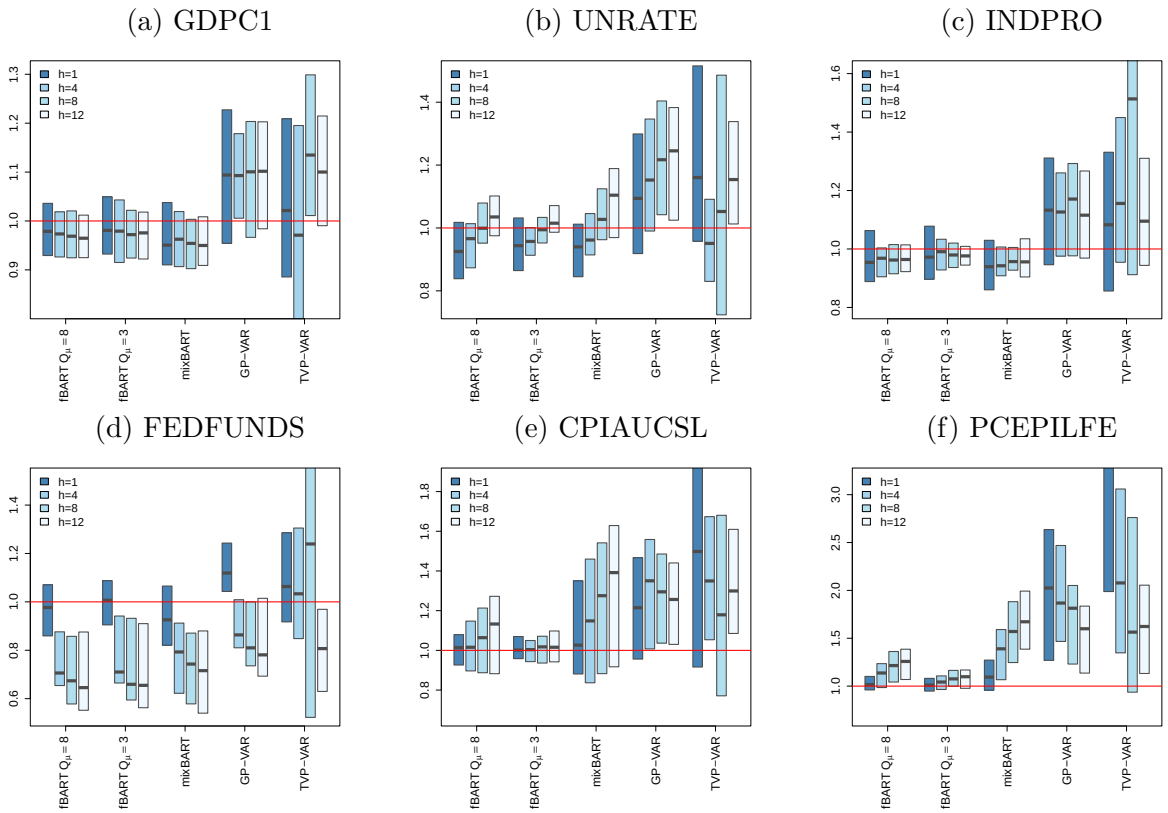


Figure 5: Boxplots of relative CRPSs to the Bayesian VAR across variables and forecast horizons.

The first subsection below starts with some reduced-form results on the importance of the factors and how the observed variables vary nonlinearly with the covariates. The second subsection provides further information on how we identify the structural shocks and presents the impulse response functions for shocks to financial conditions and monetary policy.

5.1 Reduced-form results

To provide a sense of the magnitudes of nonlinearities, Figure 6 reports the shares of variance in each series attributable to the nonlinear factor components of the model, with black lines denoting the posterior median and the upper/lower bounds being the 25th and 75th posterior percentiles. Overall, these estimates indicate the nonlinear components to be important, to a degree that varies across variables. Averaged across all series, the nonlinear factor components account for about 37% of variation. Based on posterior medians, the share is relatively high (above 50%) for series including the unemployment rate, federal tax receipts, CPI inflation, compensation growth, job growth, and housing starts. The share is relatively low for growth in GDP, productivity, TFP,

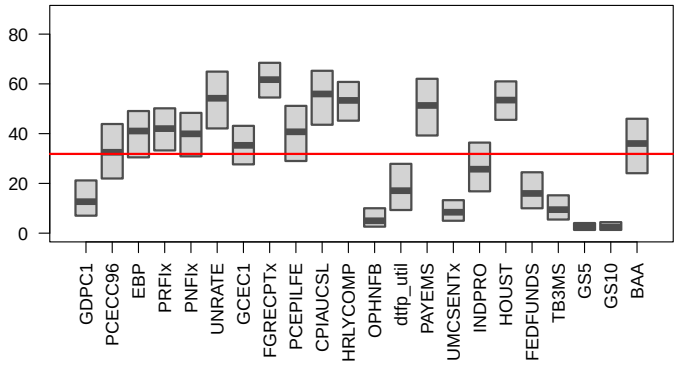
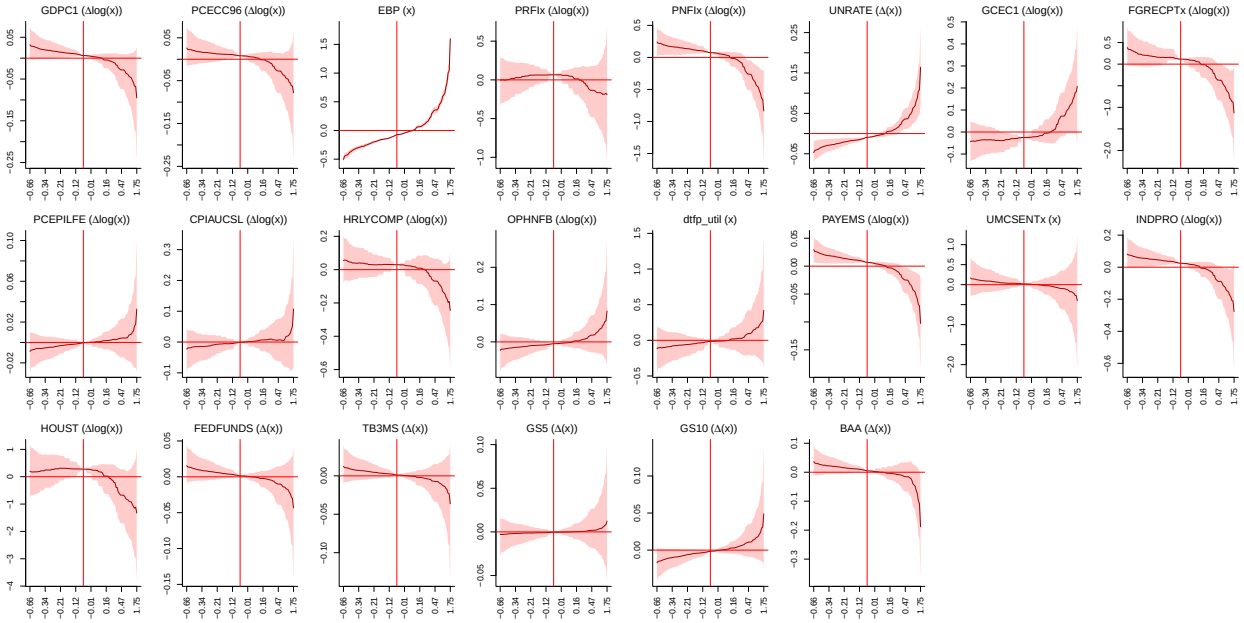


Figure 6: Variation explained (in %) of the nonlinear factor

consumer sentiment, and interest rates.



Notes: The figure shows how the observed series change with the EBP relative to a benchmark scenario where all variables are fixed at the median. Red areas refer to the 16^{th} and 84^{th} percentiles of the posterior distribution while the solid dark red line is the posterior median.

Figure 7: Sensitivity of the observed series to changes in the EBP

To shed some light on the role of financial conditions in macroeconomic developments, Figure 7 shows posterior distributions of how the variables of the model change with movements in the lagged EBP (keeping the remaining elements in $\boldsymbol{x}_t$ at the sample median) while subtracting the mean prediction arising from fixing all elements of $\boldsymbol{x}_t$ at the sample median. Movements in the EBP impact other variables through both the linear VAR component and the nonlinear function component $\boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t)$. The relationships of other variables to the EBP seem largely linear across

much of the distribution, but show nonlinearities with high values of the bond premium. The linear relationship detected for values below the 80% quantile of the EBP differ in strength. The nonlinearity that emerges with right tail values of the EBP is quite sharp for variables such as business investment, the unemployment rate, payroll employment, and long-term bond yields. Most other variables, such as GDP growth, also show some nonlinearity, while the impacts are relatively muted for a few variables, such as consumer sentiment.

5.2 Nonlinear responses to identified monetary policy shocks

Turning to impulse response estimates for monetary policy shocks, as noted above we use sign restrictions to identify shocks to demand, supply, and monetary policy, and a financial shock. Sign restrictions are implemented in the tradition of Baumeister and Hamilton (2015) and Korobilis (2022) through priors truncated to the respective set implied by the sign restrictions. In our framework, this implies changing the Gaussian prior on $\mathbf{\Lambda}_q$ to a truncated Gaussian prior to attach an economic meaning to the elements in $\boldsymbol{q}_t$ (and thus avoid identification issues related to column switching). We achieve this by setting:

$$\lambda_{q,ij} \sim \begin{cases} \mathcal{N}_{0,\infty}(0, 10^2), & \text{if shock } j \text{ is restricted to have a positive impact effect on variable } i, \\ \mathcal{N}_{-\infty,0}(0, 10^2), & \text{if shock } j \text{ is restricted to have a negative impact effect on variable } i, \\ d_c(\lambda_{q,ij}), & \text{if shock } j \text{ is restricted to have a fixed impact effect } c \text{ on variable } i, \\ \mathcal{N}(0, 10^2), & \text{if shock } j \text{ exerts an unrestricted impact effect on variable } i. \end{cases}$$

Here, we let $d_c(\lambda_{ij})$ denote the Dirac delta function, which puts point mass on c. If we use a truncated prior, the posterior simulation step for $\mathbf{\Lambda}_q$ changes slightly. Instead of simulating the rows of $\mathbf{\Lambda}_q$ from unrestricted Gaussian distributions, the truncated prior implies that the posterior is truncated to the set implied by the prior.

Table 4 lists the sign restrictions. For example, following typical practice, a one standard deviation monetary policy shock (tightening) is identified as a shock that raises the federal funds rate and reduces GDP growth, increases the unemployment rate, and reduces core PCE inflation. A shock to financial conditions is identified as a shock that raises the EBP and reduces GDP growth, consumption, investment, consumer sentiment, and core PCE inflation, and raises the unemployment rate, leading to a reduction of the federal funds rate. These identifying restrictions are similar to those used in studies such as Chan, Matthes, and Yu (2025).

In the interest of brevity, we focus on financial and monetary policy shocks. To assess possible

Shock	GDPC1	PCECC96	EBP	PRFIx	PNFIx	UNRATE	GCEC1	FGRECPTx	PCEPILFE	CPIAUCSL	HRLYCOMP	OPHNFB	dtfp_util	PAYEMS	UMCSENTx	INDPRO	HOUST	FEDFUNDS	TB3MS	GS5	GS10	BAA
Demand	+				-	-		+	+	+			0		+	+		+	+			
Monetary	-					+			-	-			0	-		-		+	+	+	+	+
Supply	-	-				+			+	+												
Financial	-	-	+		-	+			-						-			-				

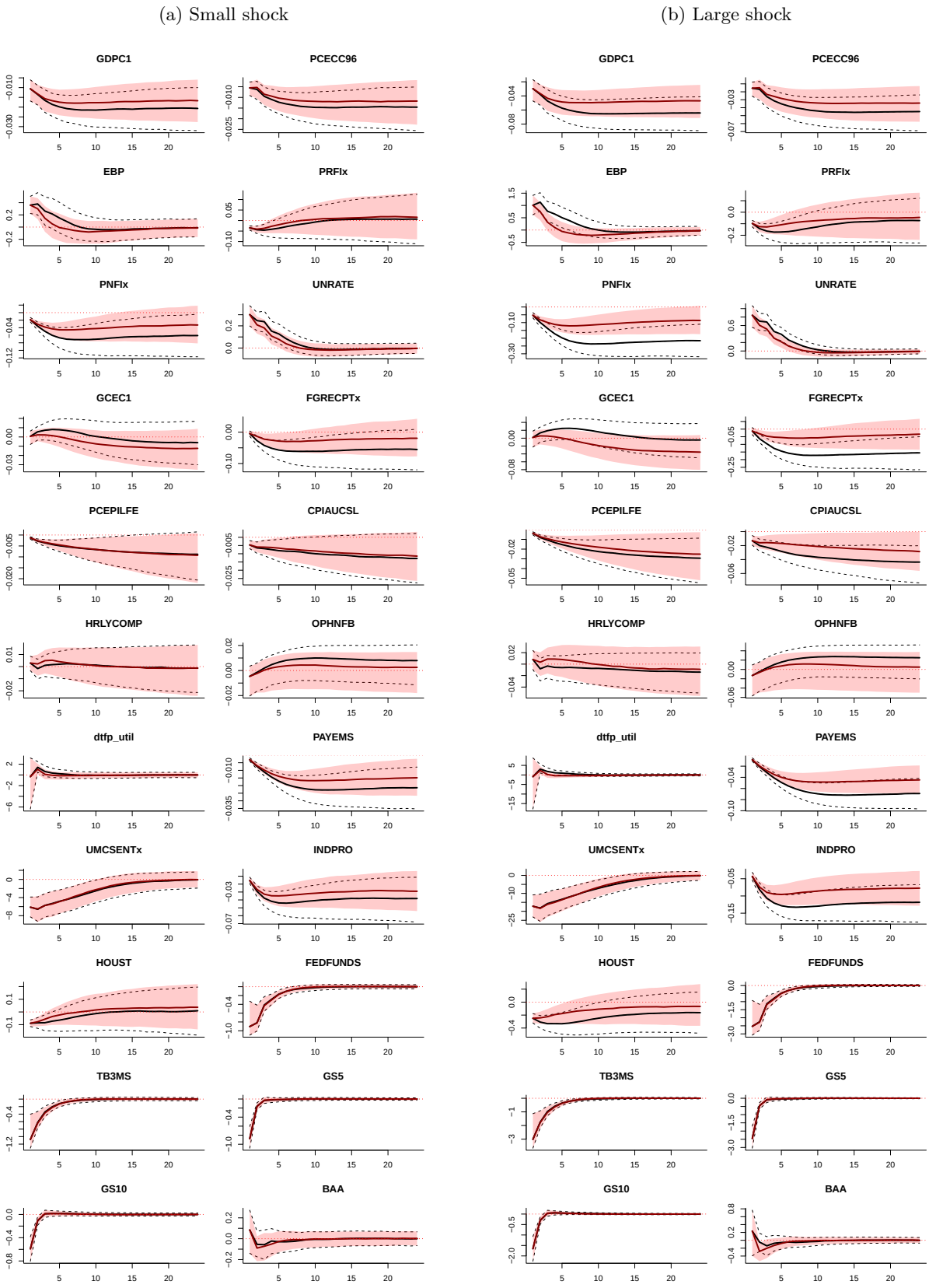
Notes: '+' refers to a positive impact reaction, '-' refers to a negative one, and 'c' implies that we fix the impact effect, whereas no sign means that we have introduced no restriction.

Table 4: Sign restrictions used to identify the structural factors in the model

nonlinearities, we consider the impacts of shocks differing in sign and size. Specifically, we estimate impulse response functions (IRFs) to positive and negative shocks, with the responses to the negative shock multiplied by -1 to facilitate comparing responses to positive and negative shocks. We consider both small and large shocks. The shocks are scaled so that the monetary policy shock (in the tightening case) triggers a median increase in FEDFUNDS of about 20 basis points in the small shock case and about 55 basis points in the large shock case. The financial shock causes an on-impact increase in the EBP of about 0.35 in the small case and 1.0 in the large case. In all cases, we compute the impulse response functions (IRFs) using the generalized IRF approach of Koop, Pesaran, and Potter (1996).

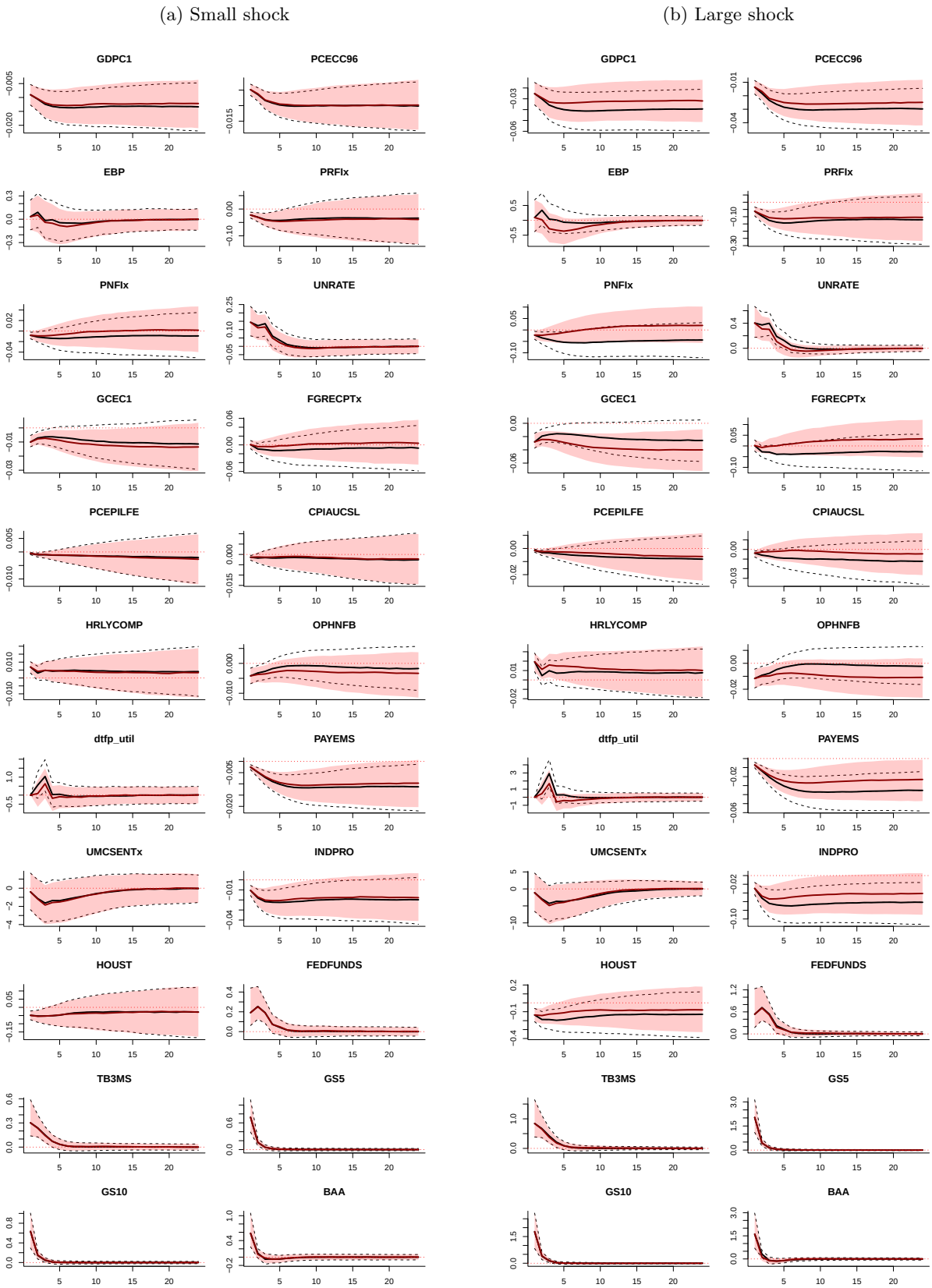
Starting with financial shocks, Figure 8 reports the posterior median responses for positive and negative shocks, along with 68% credible sets. The left- and right-side panels provide responses to small and large shocks, respectively. The estimates in panel (a) show that, with a small shock, responses to positive (tightening) and negative (easing) shocks to financial conditions are typically symmetric. More restrictive financial conditions slow economic activity (e.g., reducing GDP, consumption, investment, industrial production, housing starts, and employment, while raising unemployment), whereas an easing has the reverse impact (with sign flipped) and very similar magnitudes. The same applies to interest rates, including the 3-month T-bill rate and longer-term yields. Shocks to financial conditions also lead to small reductions in the price level, with symmetry for responses to positive and negative shocks. Broadly, our estimated impacts of shocks to financial conditions align with those obtained by Korobilis (2022).

As indicated in panel (b), large shocks have more clearly asymmetric effects in tightenings versus easings of financial conditions. With a large shock, a significant tightening of financial



Notes: The figure shows the 16th and 84th posterior percentiles and median impulse responses to positive (contractionary, dashed and solid black lines) and negative (expansionary, red shaded areas) financial shocks. To facilitate comparability we multiply the IRFs to negative shocks by -1. The numbers on top of the IRFs correspond to the posterior probability that the reactions to a positive shock are stronger than the median reactions to a negative shock.

Figure 8: Impulse responses to positive and negative financial shocks of different sizes



Notes: The figure shows the 16th and 84th posterior percentiles and median impulse responses to positive (contractionary, dashed and solid black lines) and negative (expansionary, red shaded areas) monetary policy shocks. To facilitate comparability we multiply the IRFs to negative shocks by -1. The numbers on top of the IRFs correspond to the posterior probability that the reactions to a positive shock are stronger than the median reactions to a negative shock.

Figure 9: Impulse responses to positive and negative monetary policy shocks of different sizes

conditions has bigger impacts on economic activity and inflation than does a corresponding easing of policy. By roughly six to eight quarters after the shock, compared to the small shock the large shock induces notably larger declines in economic activity measures including GDP, consumption, investment, industrial production, payroll employment, and federal tax receipts, along with a larger rise in unemployment. The large shock also induces a decline in the price level, with modestly more change following a tightening than an easing. The responses of consumer sentiment and interest rates appear to be largely symmetric for negative and positive shocks.

Figure 9 provides estimated responses to monetary policy shocks. As indicated in panel (a), with a small shock, there appears to be little asymmetry in responses to tightenings versus easings of monetary policy; results are very similar in absolute value. A policy tightening slows economic activity (e.g., reducing GDP growth, consumption, investment, industrial production, housing starts, and employment, while raising unemployment), whereas an easing has the reverse impact (with sign flipped) and essentially the same magnitudes. The policy tightening also induces declines in prices as measured with the headline CPI and core PCE price index, with symmetric effects from policy easings.

As indicated in panel (b), a large shock yields more evidence of asymmetries in responses to tightenings versus easings of monetary policy. With a large shock, a significant tightening of monetary policy has bigger impacts on economic activity and inflation than does a corresponding easing of policy. By roughly six to eight quarters after the shock, compared to the small shock the large shock induces notably larger declines in economic activity measures including GDP, investment, industrial production, and payroll employment. The large shock also induces a decline in the price level, with modestly more change following a tightening than an easing. On the other hand, the responses of a handful of variables, including consumer sentiment and interest rates, appear to be largely symmetric for negative and positive shocks.

6 Concluding remarks

We have proposed an extension to the standard VAR where departures from linearity are modeled using a nonlinear factor structure. The nonlinear factor is modeled using Bayesian regression tree methods. The use of a factor structure ensures parsimony and is motivated by the empirical findings that macroeconomic time series are typically found to be driven by a small number of

factors. The fact that these factors are nonlinear and estimated nonparametrically reduces the risk of misspecification. These features are of particular use for the researcher who expects their dataset to exhibit nonlinearities but has little guidance as to the precise form these nonlinearities will take. In an artificial data exercise and an empirical study involving US macroeconomic data, we demonstrate the effectiveness of our methods and their ability to pick out patterns in the data that cannot be found using linear methods.

It would be straightforward to extend our model to allow for other features of the VAR to be modeled more flexibly. For instance, the assumption of Gaussian errors could be easily relaxed through a mixture of Gaussians. Or the error variances could be allowed to be time varying using BART methods as in Pratola, et al. (2020). But our factor-BART VAR already exhibits a high degree of flexibility.

References

Bai, Jushan, and Serena Ng (2007), "Determining the number of primitive shocks in factor models," *Journal of Business & Economic Statistics*, 25, 52–60, https://doi.org/10.1198/073500106000000413.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010), "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92, https://doi.org/10.1002/jae.1137.

Baumeister, Christiane, Pascal Frank, Florian Huber, and Gary Koop (2025), "Oil, inflation expectations, and household characteristics: A nonlinear heterogeneous agent VAR approach," Mimeo, University of Notre Dame and University of Salzburg and University of Strathclyde.

Baumeister, Christiane, and James D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information," *Econometrica*, 83, 1963–1999, https://doi.org/10.3982/ECTA12356.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2005), "Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120, 387–422, https://doi.org/10.1162/0033553053327452.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019), "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154, https://doi.org/10.1016/j.jeconom.2019.04.024.

Carvalho, Carlos M, Nicholas G Polson, and James G Scott (2009), "Handling sparsity via the horseshoe," in *Artificial Intelligence and Statistics*, 73–80, PMLR.

Chan, Joshua C. C., Christian Matthes, and Xuewen Yu (2025), "Large structural VARs with multiple sign and ranking restrictions," March, manuscript.

Chan, Joshua C.C., Eric Eisenstat, and Rodney W. Strachan (2020), "Reducing the state space dimension in a large TVP-VAR," *Journal of Econometrics*, 218, 105–118, https://doi.org/10.1016/j.jeconom.2019.11.006.

Chernis, Tony, Niko Hauzenberger, Haroon Mumtaz, and Michael Pfarrhofer (2025), "A Bayesian Gaussian process dynamic factor model."

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (2010), "BART: Bayesian additive regression trees," *The Annals of Applied Statistics*, 4, 266–298, https://doi.org/10.1214/09-AOAS285.

Clark, Todd E, Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer (2023), "Tail forecasting with multivariate Bayesian additive regression trees," *International Economic Review*, 64, 979–1022, https://doi.org/10.1111/iere.12619.

Cogley, Timothy, and Thomas J Sargent (2005), "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302, https://doi.org/10.1016/j.red.2004.10.009.

Crump, Richard K., Stefano Eusepi, Domenico Giannone, Eric Qian, and Argia Sbordone (2025), "A large Bayesian VAR of the United States economy," *International Journal of Central Banking*, 21, 351–209.

D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101, https://doi.org/10.1002/jae.1257.

Farrell, Max H, Tengyuan Liang, and Sanjog Misra (2021), "Deep neural networks for estimation and inference," *Econometrica*, 89, 181–213, https://doi.org/10.3982/ECTA16901.

Favara, Giovanni, Simon Gilchrist, Kurt F. Lewis, and Egon Zakrajšek (2016), "Recession Risk and the Excess Bond Premium," https://doi.org/10.17016/2380-7172.1739, FEDS Notes. Washington: Board of Governors of the Federal Reserve System, April 8, 2016.

Fernald, John (2014), "A quarterly, utilization-adjusted series on total factor productivity," https://doi.org/10.24148/wp2012-19.

Forni, Mario, Luca Gambetti, Antonio Granese, Luca Sala, and Stefano Soccorsi (2025), "An American macroeconomic picture: Supply and demand shocks in the frequency domain," *American Economic Journal: Macroeconomics*, 17, p. 311–341, https://doi.org/10.1257/mac.20230295.

Fox, Emily B, and David B Dunson (2015), "Bayesian nonparametric covariance regression," *The Journal of Machine Learning Research*, 16, 2501–2542.

Gilchrist, Simon, and Egon Zakrajšek (2012), "Credit spreads and business cycle fluctuations," *American Economic Review*, 102, 1692–1720, https://doi.org/10.1257/aer.102.4.1692.

Gonçalves, Sílvia, Ana María Herrera, Lutz Kilian, and Elena Pesavento (2024), "Nonparametric local projections," *Federal Reserve Bank of Dallas Working Paper*, 2414, https://doi.org/10.24149/wp2414.

Guerrón-Quintana, Pablo, Alexey Khazanov, and Molin Zhong (2023), "Financial and macroeconomic data through the lens of a nonlinear dynamic factor model," *FEDS working paper*, https://doi.org/10.17016/FEDS.2023.027.

Hauzenberger, Niko, Florian Huber, Karin Klieber, and Massimiliano Marcellino (2025a), "Machine learning the macroeconomic effects of financial shocks," *Economics Letters*, 250, p. 112260, https://doi.org/10.1016/j.econlet.2025.112260.

Hauzenberger, Niko, Florian Huber, Massimiliano Marcellino, and Nico Petz (2025b), "Gaussian process vector autoregressions and macroeconomic uncertainty," *Journal of Business & Economic Statistics*, 43, 27–43, https://doi.org/10.1080/07350015.2024.2322089.

Huber, Florian, and Martin Feldkircher (2019), "Adaptive shrinkage in Bayesian vector autoregressive models," *Journal of Business & Economic Statistics*, 37, 27–39, https://doi.org/10.1080/07350015.2016.1256217.

Huber, Florian, and Gary Koop (2024), "Fast and order-invariant inference in Bayesian VARs with nonparametric shocks," *Journal of Applied Econometrics*, 39, 1301–1320, https://doi.org/10.1002/jae.3087.

Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2023), "Nowcasting in a pandemic using non-parametric mixed frequency VARs," *Journal of Econometrics*, 232, 52–69, https://doi.org/10.1016/j.jeconom.2020.11.006.

Huber, Florian, and Luca Rossini (2022), "Inference in Bayesian additive vector autoregressive tree models," *The Annals of Applied Statistics*, 16, 104–123, https://doi.org/10.1214/21-AOAS1488.

Kastner, Gregor (2019), "Sparse Bayesian time-varying covariance estimation in many dimensions," *Journal of Econometrics*, 210, 98–115, https://doi.org/10.1016/j.jeconom.2018.11.007.

Kastner, Gregor, Sylvia Frühwirth-Schnatter, and Hedibert Freitas Lopes (2017), "Efficient Bayesian inference for multivariate factor stochastic volatility models," *Journal of Computational and Graphical Statistics*, 26, 905–917, https://doi.org/10.1080/10618600.2017.1322091.

Kolesár, Michal, and Mikkel Plagborg-Møller (2025), "Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly," *Journal of Business & Economic Statistics*, 43, 737–754, https://doi.org/10.1080/07350015.2025.2539478.

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

Korobilis, Dimitris (2022), "A new algorithm for structural restrictions in Bayesian vector autoregressions," *European Economic Review*, 148, p. 104241, https://doi.org/10.1016/j.euroecorev.2022.104241.

Linero, Antonio R. (2025), "Defensive model expansion for robust Bayesian inference," *https://arxiv.org/abs/2510.09598*.

Makalic, Enes, and Daniel F. Schmidt (2015), "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23, 179–182, https://doi.org/10.1109/LSP.2015.2503725.

McCracken, Michael W., and Serena Ng (2021), "FRED-QD: A Quarterly Database for Macroeconomic Research," *Federal Reserve Bank of St. Louis Review*, 103, 1–44, https://doi.org/10.20955/r.103.1-44, First Quarter.

Müller, Ulrich K., and Mark W. Watson (2025), "Forecasting related time series," *Journal of Applied Econometrics*, forthcoming, https://doi.org/10.1002/jae.70050.

Pratola, Matthew T., Hugh A. Chipman, Edward I. George, and Robert E. McCulloch (2020), "Heteroscedastic BART via multiplicative regression trees," *Journal of Computational and Graphical Statistics*, 29, 405–417, https://doi.org/10.1080/10618600.2019.1677243.

Primiceri, Giorgio E (2005), "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72, 821–852, https://doi.org/10.1111/j.1467-937x.2005.00353.x.

Shin, Minsuk, Anirban Bhattacharya, and Valen E. Johnson (2020), "Functional horseshoe priors for subspace shrinkage," *Journal of the American Statistical Association*, 115, 1784–1797, https://doi.org/10.1080/01621459.2019.1654875.

Sims, Christopher A, and Tao Zha (2006), "Were there regime switches in US monetary policy?" *American Economic Review*, 96, 54–81, https://doi.org/10.1257/000282806776157678.

Tang, Tao, Simon Mak, and David Dunson (2024), "Hierarchical shrinkage Gaussian processes: applications to computer code emulation and dynamical system recovery," *SIAM/ASA Journal on Uncertainty Quantification*, 12, 1085–1112, https://doi.org/10.1137/23M1550682.

Velasco, Sofia (2025), "Let the tree decide: FABART a non-parametric factor model."

Williams, Christopher KI, and Carl Edward Rasmussen (2006), *Gaussian Processes for Machine Learning*, 2: MIT Press Cambridge, MA.

Online Appendix

A Data Appendix

Table A.1: Overview of the dataset

Variable	Description	Transformation
GDPC1	Real Gross Domestic Product	5
PCECC96	Real Personal Consumption Expenditures	5
EBP	Excess Bond Premium	1
PRFIx	Residential fixed investment	5
PNFIx	Non-residential fixed investment	5
UNRATE	Unemployment rate	2
GCEC1	Government consumption expenditures	5
FGRECPTx	Federal tax receipts	5
PCEPILFE	Core PCE price index	6
CPIAUCSL	Consumer Price Index	6
HRLYCOMP	Hourly compensation	6
OPHNFB	Labor productivity (nonfarm business)	5
dtfp_util	Utilization-adjusted total factor productivity	1
PAYEMS	Payroll employment	5
UMCSENTx	Consumer sentiment index	1
INDPRO	Industrial production	5
HOUST	Housing starts	5
FEDFUNDS	Federal funds rate	2
TB3MS	3-month Treasury bill rate	2
GS5	5-year Treasury yield	2
GS10	10-year Treasury yield	2
BAA	BAA corporate bond yield	2

Notes: The time series are sourced from the McCracken & Ng database (except EBP). The transformation codes mean: 1 = raw series, 2 = first differences, 3 = second differences, 4 = logarithm, 5 = log first differences, 6 = second differences of logs.

B Technical Appendix

B.1 Specific prior choices

We use a horseshoe prior on the VAR coefficients:[9]

$$a_{ij} \sim \mathcal{N}(\underline{a}_{ij}, \psi^2_{a,ij}\tau^2_a), \quad \psi_{a,ij} \sim \mathcal{C}^+(0,1), \quad \tau_a \sim \mathcal{C}^+(0,1).$$

Here, $\underline{a}_{ij}$ is a prior mean that can be set such that the elements in $\boldsymbol{y}_t$ are forced toward a random walk (if they are non-stationary) or toward a white noise process (if they are stationary). Note that this prior hyperparameter choice is centered over coefficient values that imply that each element of $\boldsymbol{y}_t$ is a persistent AR processes. Hence the nonlinear factor part of the model has a strong tendency to soak up deviations from the persistent part of $\boldsymbol{y}_t$.

In contrast to the prior on the factor loadings $\boldsymbol{\Lambda}_\mu$, the prior on the linear VAR part of the model features a single global shrinkage parameter and hence small values of τ_a force all elements in $\boldsymbol{A}$ to zero. The idiosyncratic scaling parameters then allow for non-zero effects by leading to a heavy-tailed marginal (of $\psi_{a,ij}$) prior on a_{ij}.

On the elements of $\boldsymbol{\Lambda}_q$, $\lambda_{q,ij}$, we use Gaussian priors with zero mean and variance 10^2. This noninformative choice implies little shrinkage on the elements of $\boldsymbol{\Lambda}_q$. In our structural analysis in Section 5 where we use sign restrictions, we modify this prior to be a truncated Gaussian distribution. For the diagonal elements in $\boldsymbol{\Omega}$ we consider conjugate inverse Gamma priors with hyperparameters $a_\omega = b_\omega = 0.01$ to render the prior weakly informative.

For the regression trees, we adopt the standard prior specification proposed in Chipman, George, and McCulloch (2010) with $S = 250$ trees. We use these standard settings because they have been shown to perform well across a wide variety of applications, particularly when modeling US macroeconomic data. In addition, in Huber, et al. (2023) we conduct robustness checks and find that deviations from this baseline specification have only a minor impact on predictive accuracy.

This prior involves three aspects. First, the probability that a given node is non-terminal decreases in the depth d of the node: $\mathfrak{a}(1+d)^{-\mathfrak{b}}$ with $\mathfrak{a} = 0.95$ and $\mathfrak{b} = 2$. Second, the splitting

[9] The prior on the VAR's coefficients could alternatively take a Minnesota form. For VAR forecast accuracy, global-local priors such as the horseshoe often perform comparably to slightly better than the Minnesota prior (e.g., Huber and Feldkircher (2019)).

variable assignments are uniformly distributed so that, a priori, each splitting variable is equally likely. Third, conditional on the chosen splitting variable, we use a uniform prior on the discrete set of available splitting values.

On the terminal node parameters $\mu_{ij,s}$ we use a Gaussian prior. We follow Chipman, George, and McCulloch (2010) and normalize each factor to lie between -0.5 and 0.5. Then:

$$\mu_{ij,s} \sim \mathcal{N}\left(0, \frac{0.5}{S\sqrt{\mathfrak{m}}}\right),$$

where $\mathfrak{m} = 2$ is a variance hyperparameter. Larger values of S reduce the prior variance, so that each tree acts as a weak learner and the risk of over-fitting is controlled.

B.2 Details on the full conditional posterior distributions

The following section provides the full conditional posterior distributions required for posterior simulation of the factor-BART VAR. The joint posterior is intractable but, given that the priors are conditionally conjugate, we can implement a Gibbs sampler that cycles through the following steps.

- **Step I — VAR coefficients $\boldsymbol{A}$:** Let $\boldsymbol{a}_s'$ denote the s^{th} row of $\boldsymbol{A}$. Conditional on all other parameters, define the partially adjusted dependent variable $\boldsymbol{Y}^{(s)} = \boldsymbol{Y}_s - \boldsymbol{M}\boldsymbol{\lambda}_{\mu,s} - \boldsymbol{Q}\boldsymbol{\lambda}_{q,s}$, where $\boldsymbol{Y}_s$ is the $T \times 1$ vector stacking y_{st} and $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_T)'$ is the $T \times Q_q$ matrix of structural factors. Then:

$$\boldsymbol{a}_s \mid \bullet \sim \mathcal{N}(\bar{\boldsymbol{a}}_s, \bar{\boldsymbol{V}}_{a,s}),$$

$$\bar{\boldsymbol{V}}_{a,s} = \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{\omega_s^2} + \boldsymbol{D}_{a,s}^{-1}\right)^{-1}, \qquad \bar{\boldsymbol{a}}_s = \bar{\boldsymbol{V}}_{a,s}\left(\frac{\boldsymbol{X}'\boldsymbol{Y}^{(s)}}{\omega_s^2} + \boldsymbol{D}_{a,s}^{-1}\underline{\boldsymbol{a}}_s\right),$$

  where $\boldsymbol{X}$ is the $T \times K$ regressor matrix stacking $\boldsymbol{x}_t$, $\boldsymbol{D}_{a,s}$ is diagonal with entries $\psi_{a,sj}^2\tau_a^2$, and $\underline{\boldsymbol{a}}_s$ is the horseshoe prior mean.

- **Step II — Factor loadings $\boldsymbol{\Lambda}_\mu$:** Each row $\boldsymbol{\lambda}_{\mu,s}$ (for $s = 1, \ldots, M$) is drawn equation-by-equation from

$$\boldsymbol{\lambda}_{\mu,s} \mid \bullet \sim \mathcal{N}(\bar{\boldsymbol{\lambda}}_{\mu,s}, \bar{\boldsymbol{V}}_{\mu,s}),$$

$$\bar{\boldsymbol{V}}_{\mu,s} = \left(\frac{\boldsymbol{M}'\boldsymbol{M}}{\omega_s^2} + \boldsymbol{D}_{\mu,s}^{-1}\right)^{-1}, \qquad \bar{\boldsymbol{\lambda}}_{\mu,s} = \bar{\boldsymbol{V}}_{\mu,s}\frac{\boldsymbol{M}'}{\omega_s^2}(\boldsymbol{Y}_s - \boldsymbol{X}\boldsymbol{a}_s - \boldsymbol{Q}\boldsymbol{\lambda}_{q,s}),$$

where $\boldsymbol{D}_{\mu,s}$ is diagonal with entries $\tilde{\psi}^2_{\mu,sj} = \psi^2_{\mu,sj}\tau^2_{\mu,s}\varpi_j$.

- **Step III — Nonlinear factors $\boldsymbol{\mu}(\boldsymbol{x}_t)$ via weighted Bayesian backfitting:** Each factor function μ_j is updated in turn using BART (Chipman, George, and McCulloch, 2010). For factor j, define the residual excluding the contribution of factor j:

$$\boldsymbol{R}_j = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{A} - \sum_{k \neq j} \boldsymbol{\mu}_k \boldsymbol{\lambda}'_{\mu,k},$$

where $\boldsymbol{\mu}_k$ is the $T \times 1$ vector $(\mu_k(\boldsymbol{x}_1), \ldots, \mu_k(\boldsymbol{x}_T))'$ and $\boldsymbol{\lambda}_{\mu,k}$ is the k^{th} column of $\boldsymbol{\Lambda}_\mu$. The factor μ_j enters each equation through the scalar loading $\lambda_{\mu,ij}$. To reduce this M-dimensional regression to a univariate target, we project $\boldsymbol{R}_j$ onto $\boldsymbol{\lambda}_{\mu,j}$:

$$\tilde{r}_{jt} = \boldsymbol{\lambda}'_{\mu,j}\boldsymbol{R}_j(t) \Big/ \|\boldsymbol{\lambda}_{\mu,j}\|^2,$$

where $\boldsymbol{R}_j(t)$ denotes the t^{th} row of $\boldsymbol{R}_j$. This scalar response satisfies

$$\tilde{r}_{jt} = \mu_j(\boldsymbol{x}_t) + \tilde{\varepsilon}_{jt},$$

with time-varying variances $\varsigma_{jt} = \left(\boldsymbol{\lambda}'_{\mu,j}\boldsymbol{H}_t\boldsymbol{\lambda}_{\mu,j}\right)/\|\boldsymbol{\lambda}_{\mu,j}\|^2$, where $\boldsymbol{H}_t = \text{diag}(h_{1t}, \ldots, h_{Mt})$. The BART sampler for μ_j is then run as a standard weighted BART regression of $\tilde{r}_{jt}$ on $\boldsymbol{x}_t$ with weights w_{jt}. Within this step, the sum-of-trees representation of μ_j is updated via Bayesian backfitting, cycling through each tree and sampling the tree structure $\mathcal{T}_{j,s}$ and its terminal node parameters $\boldsymbol{m}_{j,s}$ from their full conditionals using the BART MCMC steps of Chipman, George, and McCulloch (2010).

- **Step IV — Factor loadings $\boldsymbol{\Lambda}_q$:** Each row $\boldsymbol{\lambda}_{q,s}$ is drawn from

$$\boldsymbol{\lambda}_{q,s} \mid \bullet \sim \mathcal{N}(\bar{\boldsymbol{\lambda}}_{q,s},\, \bar{\boldsymbol{V}}_{q,s}),$$

$$\bar{\boldsymbol{V}}_{q,s} = \left(\frac{\boldsymbol{Q}'\boldsymbol{Q}}{\omega_s^2} + \frac{1}{100}\boldsymbol{I}_{Q_q}\right)^{-1}, \qquad \bar{\boldsymbol{\lambda}}_{q,s} = \bar{\boldsymbol{V}}_{q,s}\,\frac{\boldsymbol{Q}'}{\omega_s^2}(\boldsymbol{Y}_s - \boldsymbol{X}\boldsymbol{a}_s - \boldsymbol{M}\boldsymbol{\lambda}_{\mu,s})\,.$$

When sign restrictions are used (Section 5.2), we replace the Gaussian prior with a truncated Gaussian, so that the posterior is also truncated to the set implied by the sign restrictions.

- **Step V — Static factors $\boldsymbol{q}_t$:** The innovation factors are drawn independently for each t:

$$\boldsymbol{q}_t \mid \bullet \sim \mathcal{N}(\bar{\boldsymbol{q}}_t,\, \bar{\boldsymbol{V}}_q),$$

$$\bar{\boldsymbol{V}}_q = \left(\boldsymbol{\Lambda}_q' \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}_q + \boldsymbol{I}_{Q_q}\right)^{-1}, \qquad \bar{\boldsymbol{q}}_t = \bar{\boldsymbol{V}}_q\, \boldsymbol{\Lambda}_q' \boldsymbol{\Omega}^{-1} (\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{x}_t - \boldsymbol{\Lambda}_\mu \boldsymbol{\mu}(\boldsymbol{x}_t))\,.$$

- **Step VI — Error variances $\boldsymbol{\Omega}$:** For each $s = 1, \ldots, M$, let $\hat{y}_{st} = \boldsymbol{a}_s' \boldsymbol{x}_t + \boldsymbol{\lambda}_{\mu,s}' \boldsymbol{\mu}(\boldsymbol{x}_t) + \boldsymbol{\lambda}_{q,s}' \boldsymbol{q}_t$. Then:

$$\omega_s^2 \mid \bullet \sim \mathcal{G}^{-1}\left(a_\omega + \tfrac{T}{2},\; b_\omega + \tfrac{1}{2}\sum_{t=1}^{T} (y_{st} - \hat{y}_{st})^2\right).$$

- **Step VII — Horseshoe hyperparameters for $\boldsymbol{A}$:** The global and local shrinkage parameters τ_a, ν_a, $\{\psi_{a,ij}\}$, $\{\nu_{a,ij}\}$ are sampled using the efficient auxiliary variable sampler of Makalic and Schmidt (2015).

- **Step VIII — Horseshoe hyperparameters for $\boldsymbol{\Lambda}_\mu$:** The shrinkage parameters for $\boldsymbol{\Lambda}_\mu$ are updated row-by-row. For row i (looping over $i = 1, \ldots, M$), let $\boldsymbol{\lambda}_{\mu,i\cdot}$ denote the relevant elements of the i^{th} row. We apply the Makalic and Schmidt (2015) sampler to the scaled loadings $\lambda_{\mu,ij}/\sqrt{\varpi_j}$, obtaining updated local parameters $\psi_{\mu,ij}$ $(j = 1, \ldots, Q_\mu)$ and the row-specific global parameter $\tau_{\mu,i}$. The combined prior variance is then set to $\tilde{\psi}_{\mu,ij}^2 = \psi_{\mu,ij}^2 \tau_{\mu,i}^2 \varpi_j$.

- **Step IX — Column-shrinkage hyperparameter ϖ:** The global column-shrinkage level ϖ controls the overall strength of shrinkage on later factors. Its full conditional is derived by noting that $\lambda_{\mu,ij} \mid \psi_{\mu,ij}, \tau_{\mu,i}, \varpi \sim \mathcal{N}(0, \psi_{\mu,ij}^2 \tau_{\mu,i}^2 \varpi / j^\kappa)$. Collecting all elements of $\boldsymbol{\Lambda}_\mu$ across rows and columns and combining with the $\mathcal{G}^{-1}(a_\varpi, b_\varpi)$ prior gives

$$\varpi \mid \bullet \sim \mathcal{G}^{-1}\left(a_\varpi + \frac{MQ_\mu}{2},\; b_\varpi + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{Q_\mu} \frac{\lambda_{\mu,ij}^2}{j^\kappa\, \psi_{\mu,ij}^2\, \tau_{\mu,i}^2}\right),$$

  with $a_\varpi = 3$, $b_\varpi = 0.03$, and $\kappa = 2$. After drawing ϖ, the column-specific penalties are updated as $\varpi_j = \varpi / j^\kappa$, and the combined prior variances $\tilde{\psi}_{\mu,ij}^2$ are recomputed accordingly.

B.3 Mixing properties of MCMC algorithm

We find that, despite its complexity, this algorithm mixes well. We back our claim by showing two MCMC convergence diagnostics: inefficiency factors and the Raftery & Lewis diagnostic of

the total number of runs required to achieve a certain level of precision for the large US macro dataset we consider in our empirical work.[10] The quantiles are computed across equations and covariates/factors (for $\boldsymbol{A}$ and $\boldsymbol{\Lambda}_q$, respectively) and across factors and time (for $\boldsymbol{\mu}(\boldsymbol{x}_t)$). These results are based on the structural application and we run eight chains (with 10,000 draws, out of which we drop the first 5,000 as burn-in, each) in parallel. For each of these parallel runs we take every 10th draw from the 5,000 retained draws. This gives, per chain, 500 draws. Combining these draws yields 4,000 draws from which we compute our functions of interest. In all cases, inefficiency factors are well below 30 in terms of the median. Only for $\boldsymbol{\mu}(\boldsymbol{x}_t)$ we find a 90% quantile of the IF of around 53. But we view this as still acceptable given the size and complexity of the model. In terms of the Raftery & Lewis diagnostic we find that the required number of runs is often below the total number of iterations.

Quantiles	10%	25%	50%	75%	90%
			Inefficiency factors		
$\boldsymbol{A}$	1.00	1.02	1.13	1.40	1.98
$\boldsymbol{\Lambda}_q$	3.04	4.22	6.99	10.90	14.83
$\boldsymbol{\mu}(\boldsymbol{x}_t)$	6.09	7.96	12.97	30.02	53.14
			Raftery & Lewis diagnostic		
$\boldsymbol{A}$	148.00	151.00	157.00	165.00	182.00
$\boldsymbol{\Lambda}_q$	198.00	243.75	527.00	807.50	1070.00
$\boldsymbol{\mu}(\boldsymbol{x}_t)$	349.80	460.00	660.00	1158.00	1711.20

[10] We specify the parameters of this statistic along the lines suggested in Primiceri (2005): quantile=0.025, desired accuracy=0.025 and the required probability of obtaining the required accuracy p=0.95.