Forecasting with an Adaptive Nonparametric VAR

Frank C. Z. Wu[1], Joshua C. C. Chan[2], and Justin L. Tobias[3]

[1]*Department of Statistics, Purdue University*
[2]*Department of Economics, Purdue University*
[3]*Department of Economics, Purdue University*

August 2026

Abstract

We develop a Bayesian VAR in which the relationships between the endogenous variables and their lagged values are modeled nonparametrically. Specifically, the conditional mean function of each equation is specified using a generalized additive model, which facilitates the interpretation of the nonlinear dynamics. We develop an adaptive smoothing prior on the functional values, which effectively shrinks the nonparametric VAR to a standard linear VAR, while at the same time enables the data to determine the degree and shape of nonlinearity when warranted. Using a US macroeconomic dataset, we find substantial nonlinear dynamics in the macroeconomic variables. The proposed nonparametric VAR also tends to forecast better than state-of-the-art benchmarks, especially during the COVID-19 pandemic period.

1 Introduction

Following the seminal work of Sims (1980) and Doan, Litterman, and Sims (1984), Bayesian vector autoregressions (BVARs) have become the workhorse model in empirical macroeconomics and forecasting. Despite their widespread use and seeming empirical success, it has long been recognized that the linearity assumptions underlying standard BVARs might be overly restrictive given the richness and complexity inherent in macroeconomic dynamics. Subsequent work has thus extended linear VARs along a few different directions; these include Markov switching dynamics (Hamilton, 1989; Psaradakis, Ravn, and Sola, 2005), threshold VARs (Anderson and Vahid, 1998), and smooth transition VARs (Teräsvirta, 1994; Gefang and Strachan, 2009). These advances are typically parametric extensions of linear BVARs, using known functional forms to model relationships believed to be nonlinear.

More recently, there has been much interest in developing nonparametric VAR approaches that relax the linear structure entirely, allowing relationships between lagged variables and future outcomes to be described by arbitrary nonlinear functions. For example, Hauzenberger, Huber, Marcellino, and Petz (2025) developed a Gaussian Process VAR capable of modeling unusual, nonlinear patterns in conditional means and applied this method to examine asymmetric macroeconomic uncertainty shocks. Similarly, Clark, Huber, Koop, Marcellino, and Pfarrhofer (2023) introduced VAR models based on Bayesian regression trees explicitly designed to improve forecasting performance during tail-risk events like COVID-19.[1] These advances were partly motivated by the superior performance of machine learning methods, such as random forests and boosted trees, during the pandemic compared to traditional models. In essence, the extreme nonlinearities and outlier events observed in the COVID-19 data required models that could deviate significantly from linear assumptions, resulting in nonparametric models that substantially outperform linear VARs during periods of crisis. The advantage of such models, however, is typically muted under normal economic conditions.

A significant limitation of many nonparametric methods—beyond their computational intensity—is their 'black-box' nature, which makes structural interpretation challenging for macroeconomic researchers and policymakers. This lack of transparency can be particularly problematic in macroeconomics, where uncovering underlying economic mechanisms is often

[1] Early contributions to this field include Giraitis, Kapetanios, and Yates (2014), who developed kernel smoothing estimators for time-varying coefficient models and provided asymptotic theory for their performance. Unlike state-space models, they treated the VAR coefficients as unknown smooth functions of time, estimating them through localized regressions. Additionally, Kapetanios, Marcellino, and Venditti (2019) introduced a nonparametric kernel estimator for large-scale TVP-VARs.

the primary objective. Unlike certain machine learning applications, such as large language models or image processing, the consequences of incorrect predictions or spurious results in macroeconomics and financial markets can be severe. The opaque nature of these algorithms impedes our ability to pinpoint the sources of prediction errors, to anticipate when errors might occur, and to identify successful mitigation strategies. This challenge has spurred the development of more interpretable nonparametric methods and methods capable of incorporating causal inference structures within machine learning approaches for macroeconomic research.

We introduce a nonparametric Bayesian VAR model that retains an additive structure, where lag components enter the system as additive right-hand side variables. We model each lag component nonparametrically using an adaptive smoothing prior, allowing flexible modeling of various types of nonlinearities in each data series. The additive structure employed in this paper affords three significant advantages. First, it facilitates the design of a scalable and efficient Bayesian backfitting algorithm capable of handling scenarios where the number of variables and lags—and thus the number of components—is very large, as demonstrated by Wu (2024).[2] Second, the additive structure enhances interpretability and allows for focused consideration of each nonparametric component. Finally, the additivity assumption is a necessary condition for the identifiability of directed acyclic causal structures among the components, an important consideration highlighted by Peters, Mooij, Janzing, and Schölkopf (2014). This becomes increasingly relevant as models grow more complex and are applied to datasets of increasing size.

Beyond the additive structure, a key contribution of our modeling approach is the use of an informative hierarchical adaptive smoothing prior for each nonparametric component in the system. This method allows the data to determine the properties of the functions being modeled and to inform the appropriate degree of local smoothing. Moreover, the additive structure affords parsimony and helps to mitigate issues of overparametrization. When combined with our designed Bayesian backfitting algorithm, the approach efficiently handles BVAR models with potentially hundreds of variables. Computational complexity grows linearly with the number of variables in our setting, avoiding well-known dimensionality limitations (and curses). The smoothing prior technique employed in this paper, initially introduced by Koop and Poirier (2004), has been successfully applied across various contexts,

[2]The backfitting algorithm introduced by Buja, Hastie, and Tibshirani (1989) has attracted significant attention for estimating additive models due to its straightforward implementation and modular capability, allowing easy incorporation of additional components into the estimation process. This algorithm iteratively estimates each component while holding others fixed at their previous iteration values, continuing until convergence across all components.

as illustrated by Koop, Poirier, and Tobias (2005), Chib and Jeliazkov (2006), Koop and Tobias (2006), Tobias and Chan (2019), Wu (2024), and Tobias and Bond (2026). In this paper, we adapt and extend this approach within the BVAR framework.

The next section formally introduces our model in detail and briefly describes the benchmark model used for comparing forecast results. In Section 3, we provide an extensive discussion of our adaptive smoothing prior, detailing the step-by-step estimation procedure and implementation of the Bayesian backfitting algorithm. This is followed by Section 4, which presents full-sample estimation results using the empirical FRED-QD dataset. Section 5 compares the forecasting performance of our proposed model against benchmark models widely used in the literature. Finally, Section 6 concludes the paper.

2 The Model Setup

To introduce the proposed additive VAR, first consider a standard VAR with n variables and p lags, and the observations are recorded over time periods $t = 1, \ldots, T$:

$$\boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\xi}_t,$$

where $\boldsymbol{a}_0$ is an $n \times 1$ vector of intercepts and $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p$ are $n \times n$ coefficient matrices. Let $\boldsymbol{x}_t = (1, \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}')'$ denote the $(np+1)$-vector of intercept and lagged values and let $\boldsymbol{\beta}_i = (a_0^i, \boldsymbol{A}_{1,i}, \ldots, \boldsymbol{A}_{p,i})'$ represent the VAR coefficients of the i-th equation. Then, we can rewrite the i-th equation of the above system succinctly as

$$y_{i,t} = \boldsymbol{x}_t' \boldsymbol{\beta}_i + \xi_{i,t}.$$

In other words, the conditional mean function for the i-th equation is simply $f(\boldsymbol{x}_t; \boldsymbol{\beta}_i) = \boldsymbol{x}_t' \boldsymbol{\beta}_i$. Instead of this linear function, we consider the following additive model:

$$y_{i,t} = a_0^i + \sum_{j=1}^{n} \sum_{k=1}^{p} f_{jk}^i(y_{j,t-k}) + \xi_{i,t}, \tag{1}$$

where $y_{j,t-k}$ is the j-th variable at lag k. Following earlier work in this area (e.g. Koop and Poirier, 2004; Koop, Poirier, and Tobias, 2005; Chib and Jeliazkov, 2006), we directly model the functional values $f_{jk}^i(y_{j,1-k}), \ldots, f_{jk}^i(y_{j,T-k})$ and specify an informative prior that reflects the belief that the function $f_{jk}^i(\cdot)$ is 'sufficiently smooth'. More specifically, we sort

$y_{j,1-k}, \ldots, y_{j,T-k}$ in ascending order to get $x_{jk}^1 < \ldots < x_{jk}^{m_{jk}}$, where m_{jk} is the number of distinct values. We then let $\boldsymbol{\theta}_{jk}^i = \left(f_{jk}^i(x_{jk}^1), \ldots, f_{jk}^i(x_{jk}^{m_{jk}})\right)'$ denote the functional values on the sorted lag values. Intuitively, if the function f_{jk}^i is 'smooth', then nearby elements in $\boldsymbol{\theta}_{jk}^i$ are expected to be close if the corresponding sorted lag values are close. We formalize this idea using an informative prior on $\boldsymbol{\theta}_{jk}^i$, and we defer the details of this prior to Section 3.1.

Let $\mathbf{y}^i = (y_1^i, \ldots, y_T^i)'$ denote the T-vector of observations of the i-th variable, and let $\boldsymbol{D}_{jk}^i$ represent the corresponding $T \times m_{jk}$ selection matrix that maps $\boldsymbol{\theta}_{jk}^i = \left(f_{jk}^i(x_{jk}^1), \ldots, f_{jk}^i(x_{jk}^{m_{jk}})\right)'$ to $(f_{jk}^i(y_{j,1-k}), \ldots, f_{jk}^i(y_{j,T-k}))'$. Note that each row of $\boldsymbol{D}_{jk}^i$ contains precisely one element that is equal to 1 and all other elements are 0; hence $\boldsymbol{D}_{jk}^i$ is sparse. Then, we can write (1) more compactly as:

$$\boldsymbol{y}^i = a_0^i \mathbf{1}_T + \sum_{j=1}^{n} \sum_{k=1}^{p} \boldsymbol{D}_{jk}^i \boldsymbol{\theta}_{jk}^i + \boldsymbol{\xi}^i, \tag{2}$$

where $\mathbf{1}_T$ is a T-vector of ones and $\boldsymbol{\xi}^i = (\xi_{i,1}, \ldots, \xi_{i,T})'$ with $\xi_{i,t}$ being the i-th element of $\boldsymbol{\xi}_t$.

To facilitate equation-by-equation estimation, we assume the reduced-form error $\boldsymbol{\xi}_t$ has a factor structure, i.e., $\boldsymbol{\xi}_t = \boldsymbol{L}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{L}$ is an $n \times r$ lower triangular matrix with ones on the main diagonal, $\boldsymbol{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with both $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \ldots, \omega_r^2)$ being diagonal. Hence, conditional on the latent factors $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T$ (but not marginal of them), we essentially have n unrelated (independent) linear regressions. Our MCMC scheme can therefore sample the functional values $\boldsymbol{\theta}_{jk}^i$ independently for each $i = 1, \ldots, n$, conditional upon the latent factors and other model parameters.

We refer to the proposed model as the adaptive smoothing additive nonparametric VAR model. Below we investigate how this model manages nonlinear dynamics and outlier phenomena observed in macroeconomic time series, and how these factors influence its forecasting performance. Additionally, we compare the forecasting performance of our specification against the Stochastic Volatility Outlier (SVO) model, which was designed specifically to accommodate outliers (Carriero, Clark, Marcellino, and Mertens (2024), hereafter referred to as 'CCMM'). The SVO model maintains a linear structure for lagged variables while augmenting the outlier volatility component in the time-varying variance-covariance matrix of VAR error terms to account for outliers.

Most macroeconomic series are inherently aggregated, and after appropriate transformations, they often appear more linear than micro-level series. For instance, total output aggregates the outputs from millions of firms, while consumption aggregates expenditures from millions of households. This aggregation process can smooth out idiosyncratic irregularities and

nonlinearities that are present at the microeconomic level. Outside exceptional periods like the COVID-19 pandemic, one might anticipate that linear models would perform as well or even better than more complex nonlinear models for many macroeconomic series, particularly those that are relatively stable and do not exhibit substantial temporal variability.[3]

In CCMM, 16 macroeconomic variables from the 'FRED-MD' and 'FRED-QD' databases, which are designed specifically for empirical macroeconomic analysis with large datasets, were selected for forecast performance comparisons among various BVAR models with different variance-covariance structures. In this paper, we select three variables from these original 16: industrial production ('INDPRO'), the S&P 500 index ('SP500') and the US dollar to UK pound spot exchange rate ('EXUSUKx'). With $p = 2$ lags, our small additive nonparametric VAR model comprises six components. Specifically, 'INDPRO' belongs to the Industrial Production series group, 'SP500' to the Stock Market series group, and 'EXUSUKx' to the Interest Rate series group. The rationale for choosing a small VAR model is our primary interest in evaluating whether a small additive nonparametric model can outperform a larger, though linear, BVAR model. If such a performance advantage is demonstrated, it would encourage the empirical adoption of smaller, computationally efficient nonparametric models. This comparison is particularly valuable given the practical advantages in terms of computational tractability, as it allows us to assess how well a compact nonparametric model performs relative to a larger parametric model with outlier-augmented adjustments.

In terms of model estimation, it is important to note that the order of these components is fixed, meaning that some variables will have their lag positions closer to the left. Additionally, among these six components, two are own lags—specifically, the $T-1$ and $T-2$ lags. Given the substantial overlap between these two lags, it is possible to combine them to reduce computational complexity. An alternative approach would be to utilize a sufficiently large grid encompassing all elements from these two lags, as employed in Tobias and Bond (2026). However, this strategy could result in an excessively large number of grid points, negatively affecting computational efficiency. For simplicity and computational ease in our subsequent forecast comparison exercises, we will use only the elements of these two lags as grid points in $\boldsymbol{x}_{jk}$. We provide detailed explanations of our model estimation procedure in the next section.

[3]Early research, such as Stock and Watson (1998), found in univariate settings that nonlinear models typically did not enhance forecasting accuracy on average. Indeed, the best-performing linear models often matched or exceeded the forecasting capabilities of nonlinear specifications across various forecast horizons and loss functions.

3 Priors and Estimation

Given the additive assumption of the conditional mean and the modular nature of MCMC methods, one can employ a backfitting algorithm to estimate the nonparametric components, i.e., each nonparametric function can be sequentially estimated while holding the remaining components fixed. In Tobias and Bond (2026), a large grid is employed, facilitating the estimation of functional values at grid points in one step, thereby integrating both estimation and prediction. For grid points not observed in the dataset, their estimated functional values can serve as predictions.

In our approach, we adjust $\boldsymbol{\theta}_{11}^{i}$ [4] so that its elements correspond to the grid points generated from the sorted values of $\mathbf{y}_T^1 = (y_0^1, \ldots, y_{T-1}^1, y_T^1)'$ rather than $\mathbf{y}_{T-1}^1$. This guarantees that we obtain the functional value at the last in–sample observation, $f_{11}^i(y_T^1)$, which is needed to form the one–step–ahead forecast y_{T+1}^1. More generally, to forecast y_{T+h}^1 we need the functional value $f_{11}^i(y_{T+h-1}^1)$. In an idealized setting, one could construct a dense, fixed grid covering the economically relevant range of the series (for example, a grid $[2.5, 2.6, \ldots, 15.0]$ for an unemployment rate that historically lies between 2.5% and 15%) and estimate functional values at all grid points in a single run. As long as y_{T+h-1}^1 lies on this grid, the corresponding $f_{11}^i(y_{T+h-1}^1)$ is available and longer–horizon forecasts can be computed recursively without further adjustments.

In practice, however, such dense grids would be prohibitively large for some macroeconomic series and would substantially increase computational cost. We therefore start from the empirical grid given by the sorted in–sample values $(y_0^1, \ldots, y_T^1)'$, which is sufficient to estimate $f_{11}^i(y_T^1)$ and hence y_{T+1}^1. After simulating y_{T+1}^1, our forecasting code checks whether this value coincides with an existing grid point. If it does, we already have an estimate of $f_{11}^i(y_{T+1}^1)$ and can proceed to compute y_{T+2}^1. If it does not, we augment the grid by adding y_{T+1}^1 and re–estimate the corresponding functional value under the adaptive smoothing prior (holding the hyperparameters fixed), thereby obtaining $f_{11}^i(y_{T+1}^1)$ and enabling the computation of y_{T+2}^1. This procedure is iterated for further horizons as needed. After specifying the parameters $\boldsymbol{\theta}_{jk}^i$ in this way, we impose the adaptive smoothing prior on these unknown functional values.

The inclusion of a factor covariance structure enables the system to function as a VAR. Since each component is modeled nonparametrically, these factors are expected to capture

[4] Similar adjustments are required for other components $\boldsymbol{\theta}_{jk}^i$; here, $\boldsymbol{\theta}_{11}^i$ is used as an illustrative example.

the comovement characteristics of macroeconomic variables. For simplicity, we utilize a classical factor structure with time-invariant factor loadings. Given our small system with only three variables, this approach is sufficient, although extending it to dynamic factors would be straightforward. Introducing factors complicates estimation slightly, but we can still estimate the additive components equation-by-equation conditional upon those factors, preserving scalability and efficiency. Once this step is completed, we estimate the factor loadings and time-varying factors following the procedure outlined in Chan, Koop, Poirier, and Tobias (2019).

3.1 An Adaptive Smoothing Prior

In this section, we outline an adaptive smoothing prior on the functional values $\boldsymbol{\theta}_{jk}^i = \left(f_{jk}^i(x_{jk}^1), \ldots, f_{jk}^i(x_{jk}^{m_{jk}})\right)'$, where $x_{jk}^1 < \ldots < x_{jk}^{m_{jk}}$ are the sorted values of $y_{j,1-k}, \ldots, y_{j,T-k}$, $j = 1, \ldots, n, k = 1, \ldots, p$. For notational convenience, below we suppress the subscript and superscript of $\boldsymbol{\theta}_{jk}^i$ and simply denote it as $\boldsymbol{\theta}$. Similarly, we write $x^l \equiv x_{jk}^l, f \equiv f_{jk}^i$ and $m \equiv m_{jk}$. Intuitively, we wish to impose the prior belief that nearby elements in $\boldsymbol{\theta}$ are 'close'. Operationally, we let $\Delta_l = x^l - x^{l-1}$ and define the following $m \times m$ matrix

$$\mathbf{H}_{\boldsymbol{\theta}} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \frac{1}{\Delta_2} & -\left(\frac{1}{\Delta_3}+\frac{1}{\Delta_2}\right) & \frac{1}{\Delta_3} & \ldots & 0 & 0 & 0 \\ 0 & \frac{1}{\Delta_3} & -\left(\frac{1}{\Delta_4}+\frac{1}{\Delta_3}\right) & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{\Delta_{m-1}} & -\left(\frac{1}{\Delta_m}+\frac{1}{\Delta_{m-1}}\right) & \frac{1}{\Delta_m} \end{pmatrix}.$$

It is easy to see that $\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta}$ is an m-vector in which the first two elements are $f(x^1)$ and $f(x^2)$, and the remaining elements are differences in pointwise slopes of the form $f'(x^l) - f'(x^{l-1})$, where $f'(x^l) = (f(x^l) - f(x^{l-1}))/(x^l - x^{l-1})$. In the limiting case where the third through m-th elements of $\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta}$ are set to zero, the nonparametric component coincides with the linear regression, with the first two elements of $\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta}$ pinning down the slope and intercept of the line. More generally, large differences in pointwise slopes imply a curve that is more wiggly, whereas small differences indicate a smoother curve. We thus set the prior mean of $\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta}$ to be $\boldsymbol{\theta}_0 = (a_1, a_2, \mathbf{0}'_{m-2})'$, where a_1 and a_2 are hyperparameters. In addition to this, we follow

Tobias and Bond (2026) to allow the prior variances of $\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta}$ to vary across elements:

$$\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta} \mid \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{V}_{\boldsymbol{\eta}}),$$

where $\boldsymbol{V}_{\boldsymbol{\eta}} = \text{diag}(\exp(\eta_1), \ldots, \exp(\eta_m))$. This prior therefore centers the nonparametric component around a regression line, while allowing it to adapt locally to deviations from linearity. Specifically, the prior variances $\{\exp(\eta_j)\}$ regulate the smoothness of the functional values and as detailed below, we assume that η_j evolves smoothly according to a random walk. Also note that the matrix $\mathbf{H}_{\boldsymbol{\theta}}$ is banded—that is, it is sparse and its non-zero elements are confined to a diagonal band. This feature can be exploited to improve computational efficiency; see for example Chib and Jeliazkov (2006).

Let $\boldsymbol{a} = (a_1, a_2)'$ and we consider a relatively non-informative normal prior for $\boldsymbol{a}$: $\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{a}}, \boldsymbol{V}_{\boldsymbol{a}})$, where $\boldsymbol{\mu}_{\boldsymbol{a}} = (0, 0)'$ and $\boldsymbol{V}_{\boldsymbol{a}} = 100 \times \mathbf{I}_2$. Preliminary results indicate that the estimates are robust to variations in $\boldsymbol{\mu}_{\boldsymbol{a}}$. If additional information about the data series is available, a more informative prior on $\boldsymbol{a}$ could be employed. To simplify computation and enhance numerical stability, we fix $a_1 = a_2 = 0$ and set the associated variances $\exp(\eta_1) = \exp(\eta_2) = 16$. This configuration serves multiple purposes in our estimation framework. First, it anchors each functional curve to zero at its left boundary, ensuring that any constant offset is absorbed by the global intercept a_0^i. This is particularly important because the global intercept in our additive model is not identifiable. Without this anchoring constraint, the model would suffer from an identification problem where shifts between the functional components and the global intercept would be indeterminate. Second, this parameterization significantly improves the numerical stability of the estimation algorithm by providing a well-defined reference point for each component. This setting is maintained consistently throughout all forecast exercises to ensure comparability of the results.

Next, the log variance parameters are assumed to follow a random walk:

$$\begin{aligned}
\eta_3 &= \eta_0 + u_3 \\
\eta_4 &= \eta_3 + u_4 \\
&\vdots \\
\eta_m &= \eta_{m-1} + u_m,
\end{aligned}$$

where η_0 is an initial condition to be estimated within the model and $u_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2), j = 3, \ldots, m$.

To estimate the volatilities $\{\eta_j\}$, the initial volatility condition η_0, and the variance σ_η^2 in the volatility equations, we utilize techniques from Omori, Chib, Shephard, and Nakajima (2007), specifically the well-known ten-component normal mixture approximation for the log chi-square(1) random variable.

We specify the prior on η_0 as $\eta_0 \sim \mathcal{N}(\mu_0, V_0)$, setting $\mu_0 = -5$ and $V_0 = 1$. Additionally, we use an inverse-gamma prior $\mathcal{IG}(a_\eta = 3, b_\eta = 5)$ on σ_η^2, as these values have been shown to facilitate more stable estimation of $\{\eta_j\}$. Stability in the volatility estimates contributes to more stable estimation of $\boldsymbol{\theta}$. Unstable estimates could otherwise cause numerical issues during the Cholesky decomposition of the posterior variance matrix for $\boldsymbol{\theta}$, particularly if some elements have disproportionately large magnitudes compared to others, potentially resulting in the variance matrix not being recognized as positive definite.

Finally, we specify normal priors on the global intercept terms a_0^i and the vector of free elements $\boldsymbol{l}_i$ in the factor loadings matrix $\boldsymbol{L}$. For σ_i^2, $i = 1, \ldots, n$, and ω_j^2, $j = 1, \ldots, r$, we follow the conventional practice by assigning independent inverse-gamma priors to each of these parameters.

3.2 Posterior Simulation

This subsection outlines the Gibbs sampling algorithm used to obtain posterior draws of all unknown parameters and latent variables. For notational simplicity and to maintain consistency with Sections 2 and 3.1, we adopt the same indices i for equations, j for variables, and k for lags. Throughout, "$\cdots$" denotes conditioning on all other parameters and data. We also use σ^2 to denote the equation-specific error variance and σ_η^2 for the variance of the random walk in the smoothing prior; these should not be confused.

Before describing the sampling steps, define the following auxiliary quantities for a given equation i:

- $\mathbf{R}_{jk}^i = \mathbf{y}^i - a_0^i \mathbf{1}_T - \sum_{(j',k') \neq (j,k)} \boldsymbol{D}_{j'k'}^i \boldsymbol{\theta}_{j'k'}^i - \boldsymbol{L}\boldsymbol{f}$, the residuals obtained by removing the contribution of all components except f_{jk}^i and the factor term;
- $\tilde{\mathbf{y}}^i = \mathbf{y}^i - \sum_{j'=1}^n \sum_{k'=1}^p \boldsymbol{D}_{j'k'}^i \boldsymbol{\theta}_{j'k'}^i - \boldsymbol{L}\boldsymbol{f}$, the residuals after removing all nonparametric components and the factor term;
- $\mathbf{e}^i = \tilde{\mathbf{y}}^i - a_0^i \mathbf{1}_T$, the residuals for sampling the intercept;

- $\mathbf{y}_{\boldsymbol{f}}^i = \mathbf{y}^i - a_0^i \mathbf{1}_T - \sum_{j'=1}^{n} \sum_{k'=1}^{p} \boldsymbol{D}_{j'k'}^i \boldsymbol{\theta}_{j'k'}^i$, the data after subtracting the intercept and all nonparametric components.

With these definitions in place, the Gibbs sampler iterates through the following steps for each equation i and each component (j, k):

Step 1. Sample $\boldsymbol{\theta}_{jk}^i$. Conditional on all other parameters, the posterior of $\boldsymbol{\theta}_{jk}^i$ is multivariate normal,

$$\big(\boldsymbol{\theta} \mid \cdots, \mathbf{y}\big) \sim \mathcal{N}\big(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{d}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\big),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \frac{\boldsymbol{D}_{jk}^{i\top} \boldsymbol{D}_{jk}^i}{\sigma_i^2} + \mathbf{H}_{\boldsymbol{\theta}}^\top \boldsymbol{V}_{\boldsymbol{\eta}}^{-1} \mathbf{H}_{\boldsymbol{\theta}}, \quad \mathbf{d}_{\boldsymbol{\theta}} = \frac{\boldsymbol{D}_{jk}^{i\top} \mathbf{R}_{jk}^i}{\sigma_i^2} + \mathbf{H}_{\boldsymbol{\theta}}^\top \boldsymbol{V}_{\boldsymbol{\eta}}^{-1} \boldsymbol{\theta}_0.$$

Step 2. Sample the mixture indicators $\boldsymbol{s}$ for $\boldsymbol{\eta}$. Following the mixture-of-normals representation of the log-chi-square distribution in Omori, Chib, Shephard, and Nakajima (2007), draw $s_3, \ldots, s_m$ independently with

$$\mathbb{P}(s_j = \ell \mid \cdots) \propto p_\ell \exp\Big\{-\frac{1}{2\omega_\ell^2}\Big[\log\big((\tilde{h}_j - \tilde{\theta}_{0,j})^2\big) - \eta_j - m_\ell\Big]^2\Big\}, \quad \ell = 1, \ldots, 10.$$

Step 3. Sample $\boldsymbol{\eta}$. We fix the first two elements, η_1 and η_2, and therefore only need to sample the remaining elements in $\boldsymbol{\eta}$, denoted as $\tilde{\boldsymbol{\eta}}$. Conditional on the indicators $\boldsymbol{s}$ and all other parameters, the vector $\tilde{\boldsymbol{\eta}}$ has a multivariate normal posterior:

$$(\tilde{\boldsymbol{\eta}} \mid \cdots, \mathbf{y}) \sim \mathcal{N}\big(\boldsymbol{\Sigma}_\eta^{-1} \mathbf{d}_\eta,\ \boldsymbol{\Sigma}_\eta^{-1}\big),$$

here

$$\boldsymbol{\Sigma}_\eta = \boldsymbol{\Omega}_{\boldsymbol{s}}^{-1} + \frac{1}{\sigma_\eta^2} \mathbf{H}_{\boldsymbol{\eta}}^\top \mathbf{H}_{\boldsymbol{\eta}}, \quad \mathbf{d}_\eta = \boldsymbol{\Omega}_{\boldsymbol{s}}^{-1}\Big[\log\big((\tilde{\mathbf{h}} - \tilde{\boldsymbol{\theta}}_0)^2\big) - \boldsymbol{m}_{\boldsymbol{s}}\Big] + \frac{1}{\sigma_\eta^2} \mathbf{H}_{\boldsymbol{\eta}}^\top \mathbf{H}_{\boldsymbol{\eta}} \boldsymbol{\mu}_{\boldsymbol{\eta}}.$$

Additional details on sampling of the indicators in the previous step and the smoothing parameters can be found in Tobias and Bond (2026).

Step 4. Sample η_0: The conditional posterior distribution of the initial smoothing parameter η_0 is also normal:

$$\big(\eta_0 \mid \cdots, \mathbf{y}\big) \sim \mathcal{N}\left(\frac{\sigma_\eta^2 V_0}{V_0 + \sigma_\eta^2}\big(\frac{\eta_3}{\sigma_\eta^2} + \frac{\mu_0}{V_0}\big),\ \frac{\sigma_\eta^2 V_0}{\sigma_\eta^2 + V_0}\right).$$

Step 5. Sample σ_η^2: The conditional posterior distribution of the log volatility variance, σ_η^2,

is inverse gamma:

$$\big(\sigma_\eta^2 \mid \cdots, \mathbf{y}\big) \sim \mathcal{IG}\left(\frac{k_{np}-2}{2} + a_\eta,\ [\frac{1}{b_\eta} + \frac{1}{2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\mu}_{\boldsymbol{\eta}})^\top \mathbf{H}_{\boldsymbol{\eta}}^\top \mathbf{H}_{\boldsymbol{\eta}} (\tilde{\boldsymbol{\eta}} - \boldsymbol{\mu}_{\boldsymbol{\eta}})]^{-1}\right).$$

Step 6. Repeat Steps 1-5 for every pair (j, k).

Step 7. Sample a_0. The global intercept has a normal conditional posterior:

$$\big(a_0 \mid \cdots, \mathbf{y}\big) \sim \mathcal{N}\big(\boldsymbol{\Sigma}_{a_0}^{-1} \mathbf{d}_{a_0}, \boldsymbol{\Sigma}_{a_0}^{-1}\big),$$

where $\boldsymbol{\Sigma}_{a_0} = V_{a_0}^{-1} + \frac{\mathbf{1}_T^\top \mathbf{1}_T}{\sigma_i^2}$ and $\mathbf{d}_{a_0} = V_{a_0}^{-1}\mu_{a_0} + \frac{\mathbf{1}_T^\top \mathbf{e}^i}{\sigma_i^2}$.

Step 8. Sample σ_i^2. The conditional posterior distribution of the error variance in the ith equation is inverse-gamma:

$$(\sigma_i^2 \mid \cdots, \mathbf{y}) \sim \mathcal{IG}\left(\frac{T}{2} + a_\sigma,\ b_\sigma + \tfrac{1}{2}(\mathbf{e}^i)^\top \mathbf{e}^i\right).$$

Step 9. Sample $\boldsymbol{f}$. To sample the factors $\boldsymbol{f}$ we follow the algorithm described in Chapter 18 of Chan, Koop, Poirier, and Tobias (2019). It follows that

$$\big(\boldsymbol{f} \mid \cdots, \mathbf{y}\big) \sim \mathcal{N}\big(\boldsymbol{\Sigma}_{\boldsymbol{f}}^{-1} \mathbf{d}_{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\boldsymbol{f}}^{-1}\big), \tag{3}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \mathbf{I}_T \otimes \boldsymbol{\Omega}^{-1} + \mathbf{I}_T \otimes (\boldsymbol{L}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{L}), \quad \mathbf{d}_{\boldsymbol{f}} = (\mathbf{I}_T \otimes (\boldsymbol{L}^\top \boldsymbol{\Sigma}^{-1}))\mathbf{y}_{\boldsymbol{f}}$.

Step 10. Sample $\boldsymbol{L}$. We sample elements of the factor loading matrix equation-by-equation; detailed steps are provided in Chan, Koop, Poirier, and Tobias (2019). The conditional posterior distributions of all these elements are normal. We omit the associated derivations here, as results directly follow from standard linear regression operations.

Step 11. Sample ω^2. These $\omega_1^2, \ldots, \omega_r^2$ are conditionally independent given the data and other parameters. Additionally, their posteriors are Inverse-gamma:

$$\big(\omega_i^2 \mid \cdots, \mathbf{y}\big) \sim \mathcal{IG}\big(\frac{T}{2} + a_{\omega_i},\ b_{\omega_i} + \frac{1}{2}\sum_{t=1}^{T} f_{it}^2\big), \quad i = 1, \ldots, r. \tag{4}$$

4 Full Sample Estimation Results

This section reports the full-sample estimation results from the proposed additive nonparametric VAR model.[5] In particular, we illustrate how the proposed model facilitates the visualization of the nonlinear, dynamic relationships among the selected macroeconomic variables. To that end, we construct a dataset using the FRED-QD database, which is a macroeconomic database designed specifically for empirical "big data" analysis. It has been extensively used in the empirical macroeconomic literature and is publicly available online at `https://www.stlouisfed.org/research/economists/mccracken/fred-databases`. Further details about this database are provided in McCracken and Ng (2020). Our sample spans from 1959Q1 to 2024Q1, comprising $T = 261$ quarterly observations.

For our empirical application, we select three variables, namely, industrial production ('INDPRO'), the S&P 500 index ('SP500') and the US dollar to UK pound spot exchange rate ('EXUSUKx'), and fit them using the proposed nonparametric VAR. Figures 1-3 display the estimated functional relationships for the three equations. Given our model specification with $p = 2$ lags, we obtain six nonparametric components for each equation: three plots correspond to the first lags of 'INDPRO', 'SP500', and 'EXUSUKx', while the remaining three plots correspond to the second lags of these variables. In each figure, we present the estimated functional values plotted against their corresponding grid point values for each nonparametric component, accompanied by the 90% point-wise credible interval bands.

These figures reveal that the dynamic relationships between many endogenous variables and their own lags or cross-variable lags exhibit regions of local nonlinearity. However, some relationships appear primarily linear, such as the impact of the first and second lags of 'INDPRO' on 'SP500'. Overall, the existence of several precisely estimated nonlinear relationships suggests the potential benefits of our small-dimensional nonparametric model for forecasting purposes. In the following section, we investigate this hypothesis explicitly by comparing the performance of the proposed model to a higher-dimensional, commonly-used Bayesian VAR.

[5] Additional estimation results from the simulation study can be found in Appendix B.

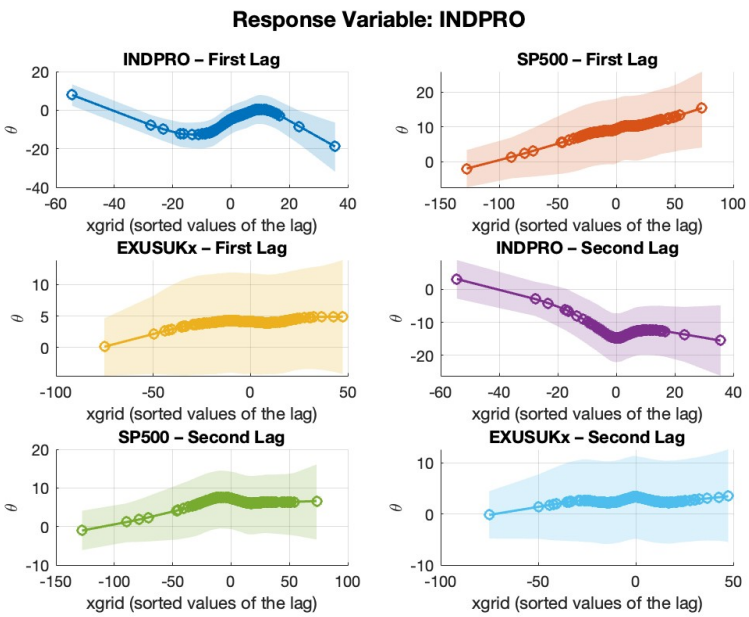

Figure 1: Plots of additive nonparametric components of the conditional mean for the 'INDPRO' equation, along with the 90% credible interval bands (point-wise).

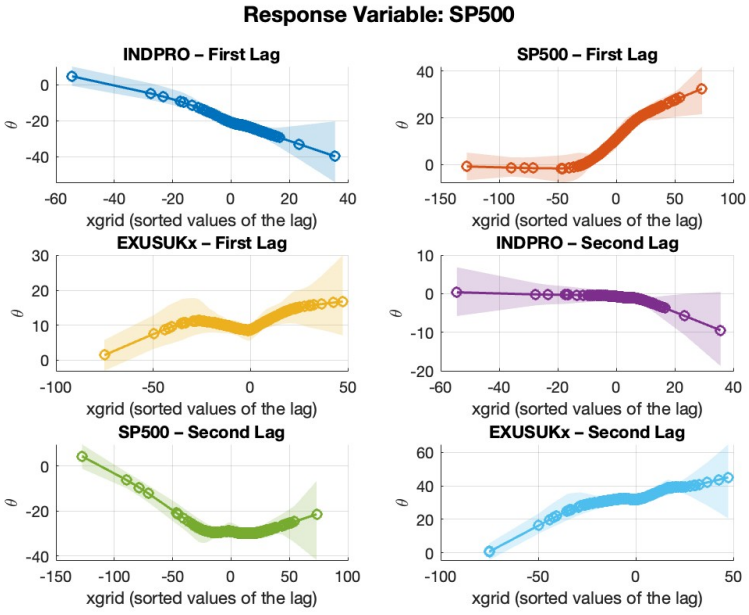

Figure 2: Plots of additive nonparametric components of the conditional mean for the 'SP500' equation, along with the 90% credible interval bands (point-wise).

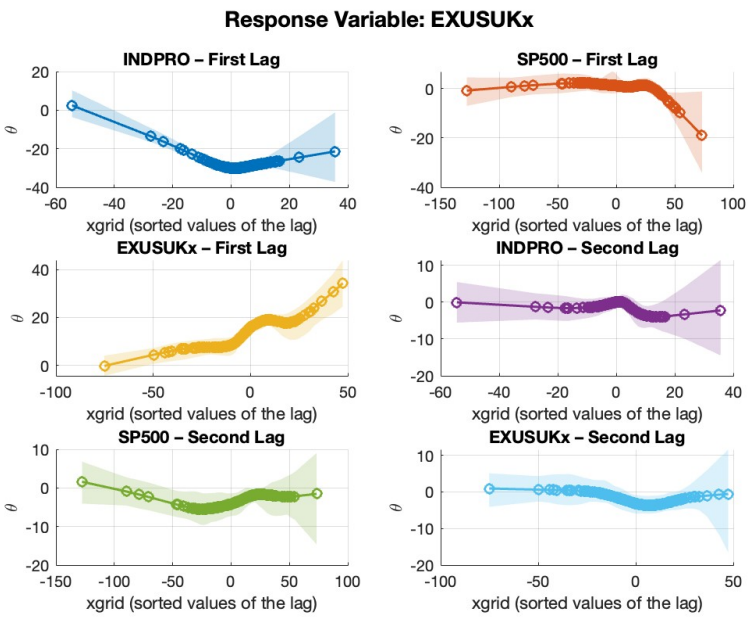

Figure 3: Plots of additive nonparametric components of the conditional mean for the 'EXUSUKx' equation, along with the 90% credible interval bands (point-wise).

Next, Figures 4–6 examine the relationship between the magnitude of the differences in slope of the nonparametric component, $|(\mathbf{H}_{\boldsymbol{\theta}}\boldsymbol{\theta})_j|$, and the corresponding variance parameter $\exp(\eta_j)$. This visualization helps assess the adaptive nature of our smoothing prior: larger slope differences should correspond to higher variances, allowing for more flexible functional forms in regions of rapid changes.

The figures confirm the expected positive relationship between the slope differences and the variance parameters. In particular, the Pearson correlations for the three equations are 0.566, 0.476, and 0.629, respectively, indicating that our adaptive smoothing prior successfully adjusts the degree of smoothness based on the local characteristics of the functional relationships.

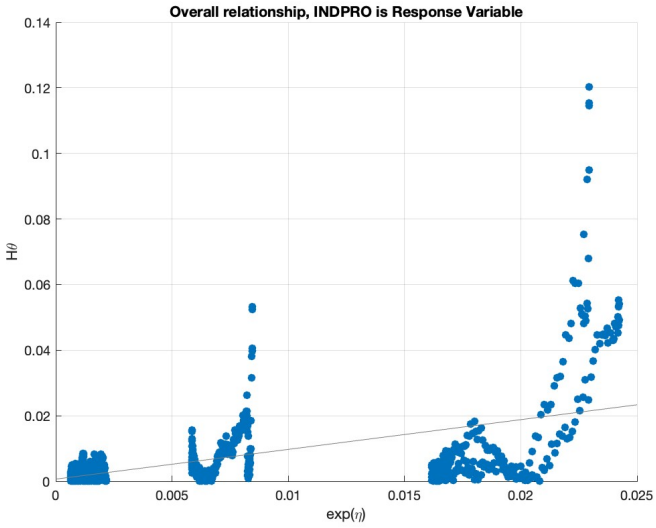

Figure 4: Absolute values of slope differences plotted against their corresponding exp($\boldsymbol{\eta}$) values for the 'INDPRO' equation.

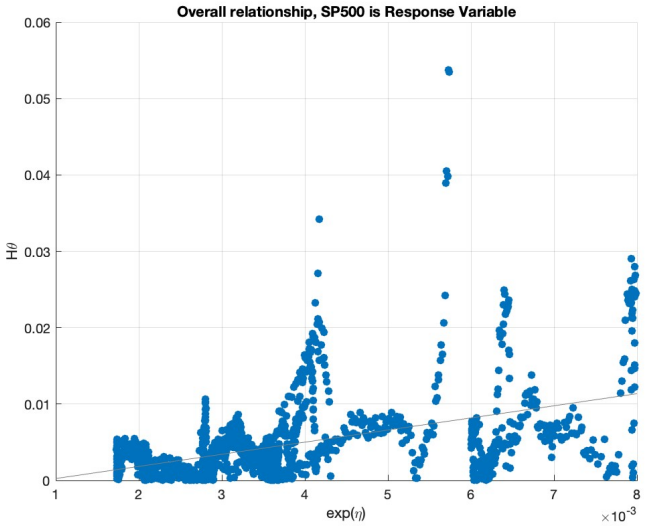

Figure 5: Absolute values of slope differences plotted against their corresponding exp($\boldsymbol{\eta}$) values for the 'SP500' equation.

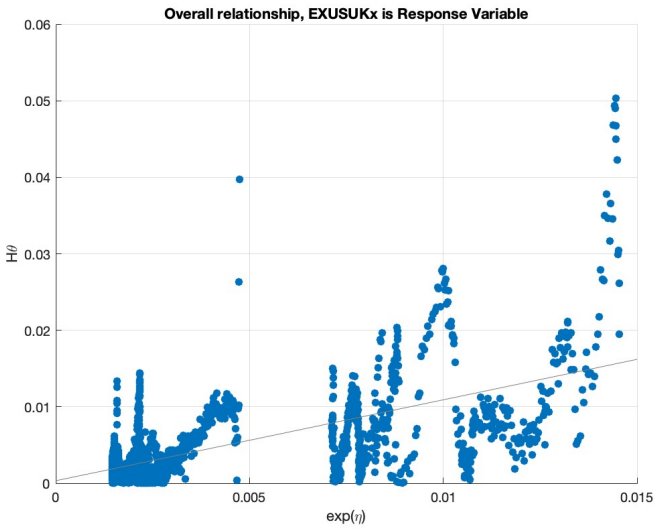

Figure 6: Absolute values of slope differences plotted against their corresponding $\exp(\boldsymbol{\eta})$ values for the 'EXUSUKx' equation.

5 Out-of-Sample Forecasting Results

In this section, we evaluate the performance of the proposed additive nonparametric VAR in an out-of-sample forecasting exercise. The benchmark model we use for comparison is the SVO model proposed by CCMM, which incorporates 16 macroeconomic variables including 'INDPRO', 'SP500', and 'EXUSUKx'. The full sample spans from 1959Q1 to 2024Q1, consisting of 261 quarterly observations, and they are all sourced from the FRED-QD dataset. To evaluate the performance of the models, we generate forecasts of the three variables, estimated using different sample sizes starting from 1959Q1. Specifically, we calculate one- to twenty-four-step-ahead forecasts and assess the model performance over different evaluation windows to identify the superior forecasting model for each particular window. These evaluation windows include: the largest window (1985Q1 to 2017Q4) for assessing overall performance, the Great Moderation period (1985Q1 to 2007Q4), the Great Financial Crisis (2008Q1 to 2014Q4), the COVID-19 period (2020Q2 to 2021Q1), and the Post-COVID-19 period (2021Q2 to 2023Q2). In this section, we highlight results for the largest window and the COVID-19 period, while additional results are provided in the Appendix A.

We evaluate point forecasts using Root Mean Squared Errors (RMSEs) and density forecasts using the Continuous Ranked Probability Scores (CRPSs), the same metrics used by CCMM

for comparing various BVAR models. According to Tibshirani (2023), CRPS is less sensitive to outliers and generally considered more robust compared to log predictive likelihoods. We calculate the CRPS using forecast draws generated from our MCMC algorithm. Tables 1-4 report the performance of the point and density forecasts from the proposed model relative to the benchmark over the specified evaluation windows. Specifically, we divide the RMSEs and CRPSs of the proposed model by those of the benchmark. Values less than one indicate that our model outperforms the benchmark. When these occur, results are highlighted in blue.

Overall, the proposed model generally outperforms the benchmark for both point and density forecasts, particularly for longer forecast horizons. The benchmark performs relatively well in terms of point forecasts of 'INDPRO' in short-horizons outside the COVID-19 evaluation window. In contrast, the proposed model consistently outperforms the benchmark for density forecasts, with the majority of the ratios below one. These results suggest that a VAR with linear conditional mean functions, even in the presence of stochastic volatility and outlier adjustments, might not adequately capture the nonlinear dynamics in some macroeconomic time series.

During the COVID-19 period, the SVO model exhibits notably weaker performance for 'INDPRO' compared to our proposed model. This contrasts sharply with results outside the COVID-19 period and strongly suggests distinctive behavior of 'INDPRO' during this period. Furthermore, as detailed in the Appendix A, our proposed model performs well in other evaluation periods, particularly for the 'SP500' and 'EXUSUKx' series.

Table 1: RMSE Comparison: Additive VAR Adaptive Smoothing vs. SVO
Evaluation Window: 1985Q1 to 2017Q4 (132 samples)
Values below 1 indicate improvement using our proposed model

Step Ahead	**INDPRO**	**SP500**	**EXUSUKx**
1	1.074	0.965	0.983
2	1.069	0.952	0.973
3	1.090	0.986	0.993
4	1.111	0.969	0.986
8	1.330	0.930	0.988
12	1.232	0.686	0.912
16	0.991	0.317	0.127
20	0.262	0.109	0.003
24	0.047	0.020	0.007

Table 2: CRPS Comparison: Additive VAR Adaptive Smoothing vs. SVO
Evaluation Window: 1985Q1 to 2017Q4 (132 samples)
Values below 1 indicate improvement using our proposed model

Step Ahead	INDPRO	SP500	EXUSUKx
1	1.077	0.896	0.973
2	0.986	0.835	0.944
3	0.931	0.801	0.918
4	0.897	0.732	0.871
8	0.849	0.541	0.734
12	0.741	0.400	0.572
16	0.560	0.282	0.424
20	0.404	0.199	0.316
24	0.296	0.142	0.220

Table 3: RMSE Comparison: Additive VAR Adaptive Smoothing vs. SVO
COVID-19 Period: 2020Q2 to 2021Q1 (4 samples)
Values below 1 indicate improvement using our proposed model

Step Ahead	INDPRO	SP500	EXUSUKx
1	1.039	1.061	0.884
2	0.914	0.925	1.691
3	0.630	1.120	1.404
4	0.166	1.581	1.057
8	0.598	0.950	0.821
12	1.244	0.914	1.278

Table 4: CRPS Comparison: Additive VAR Adaptive Smoothing vs. SVO
COVID-19 Period: 2020Q2 to 2021Q1 (4 samples)
Values below 1 indicate improvement using our proposed model

Step Ahead	INDPRO	SP500	EXUSUKx
1	1.073	0.911	0.868
2	0.812	0.897	1.452
3	0.533	0.846	0.894
4	0.202	0.801	0.636
8	0.328	0.540	0.676
12	0.428	0.266	0.380

6 Concluding Remarks

This paper developed an additive nonparametric VAR model. A smoothness prior approach was employed to estimate functions relating own lags and lags of other variables to contemporaneous outcomes of macroeconomic series. A factor structure was introduced to allow

cross-equation correlation as in a traditional VAR framework. The approach taken allowed the data to both inform the degree of smoothing of our additive functions and to determine how the extent of smoothing should change across the support of each random variable. The additive structure used in our modeling also offers several key advantages. First and foremost, compared to other nonparametric modeling approaches, our method provides greater interpretability, making it particularly suitable for analyzing macroeconomic series. Additionally, by incorporating a banded matrix within the smoothing prior and utilizing a backfitting algorithm, the approach achieves enhanced scalability and computational efficiency, which is particularly useful when analyzing large datasets with numerous variables.

Looking ahead, recent events have demonstrated that macroeconomic series can exhibit abrupt and significant changes due to a variety of sources—not only economic factors but also natural and social phenomena such as natural disasters and geopolitical tensions. Such complexities increasingly challenge policymakers' ability to forecast key macroeconomic variables accurately. This situation presents both a challenging and exciting frontier for macroeconometric research. While pursuing machine learning-based nonparametric modeling approaches to capture outlier events and nonlinearities is highly promising, designing more interpretable models that effectively handle nonlinear dynamics and uncover underlying economic mechanisms remains equally important. We believe that future research will continue to advance along both of these promising avenues.

Appendix A: Additional Forecasting Results

This appendix presents additional forecasting results for the proposed adaptive smoothing additive nonparametric VAR model and the SVO model from CCMM. Values below 1, highlighted in blue, indicate superior performance of our proposed model. The evaluation windows include the Great Moderation period, the Great Financial Crisis period, and the Post-COVID-19 period.

Table 5: RMSE Comparison: Additive VAR Adaptive Smoothing vs. SVO
Great Moderation: 1985Q1 to 2007Q4 (92 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	1.163	0.957	1.010
2	1.139	0.935	0.995
3	1.106	0.967	1.020
4	1.154	0.954	1.001
8	1.369	0.929	1.004
12	1.424	0.658	0.928
16	0.888	0.302	0.116
20	0.182	0.099	0.002
24	0.032	0.018	0.006

Table 6: CRPS Comparison: Additive VAR Adaptive Smoothing vs. SVO
Great Moderation: 1985Q1 to 2007Q4 (92 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	1.146	0.889	1.001
2	1.027	0.825	0.962
3	0.949	0.782	0.931
4	0.939	0.720	0.876
8	0.914	0.563	0.759
12	0.703	0.404	0.579
16	0.492	0.277	0.418
20	0.337	0.190	0.297
24	0.238	0.133	0.198

Table 7: RMSE Comparison: Additive VAR Adaptive Smoothing vs. SVO
Great Financial Crisis: 2008Q1 to 2014Q4 (28 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	0.901	0.972	0.891
2	1.006	0.980	0.915
3	1.113	1.019	0.976
4	1.114	0.995	0.978
8	1.108	0.918	0.905
12	1.292	0.880	0.826
16	1.462	0.337	0.765
20	1.586	0.427	0.502
24	0.926	0.308	0.294

Table 8: CRPS Comparison: Additive VAR Adaptive Smoothing vs. SVO
Great Financial Crisis: 2008Q1 to 2014Q4 (28 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	0.907	0.884	0.897
2	0.918	0.851	0.886
3	0.944	0.841	0.927
4	0.861	0.751	0.869
8	0.634	0.440	0.646
12	0.636	0.335	0.531
16	0.512	0.233	0.432
20	0.433	0.174	0.348
24	0.486	0.143	0.288

Table 9: RMSE Comparison: Additive VAR Adaptive Smoothing vs. SVO
Post-COVID-19 Period: 2021Q2 to 2023Q2 (9 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	0.563	0.832	1.118
2	0.426	0.915	0.998
3	1.115	0.985	0.916
4	2.202	1.005	0.911

Table 10: CRPS Comparison: Additive VAR Adaptive Smoothing vs. SVO
Post-COVID-19 Period: 2021Q2 to 2023Q2 (9 samples)

Step Ahead	INDPRO	SP500	EXUSUKx
1	0.528	0.730	0.982
2	0.380	0.733	0.868
3	0.415	0.720	0.836
4	0.454	0.752	0.814

Appendix B: Simulation Study

With multiple equations in our model, each equation for the i-th variable is a nonparametric additive equation with several components. In this section, we primarily focus on the estimation results of the slope coefficients of the linear components and assess whether our algorithm can distinguish between linear and nonlinear components through simulation studies by generating artificial datasets. Even with three variables and six components for each variable, we have 18 components in total in the system. Practically, we cannot run simulations on every combination of linear and nonlinear specifications for these 18 components. Additionally, it is expected that if there are a large number of complicated nonlinear components, our estimation results would not be accurate or we would need a large number of data points.

Hence, we narrow our investigation to four data generating models. These four models used to generate each $\boldsymbol{y}^i$ in our systems are described below.

The first model is fully linear with six linear components:

$$\boldsymbol{y}^i = 2 \cdot \mathbf{1}_T + \mathrm{Z}_1^i + 3\mathrm{Z}_2^i + 5\mathrm{Z}_3^i + 7\mathrm{Z}_4^i + 9\mathrm{Z}_5^i + 11\mathrm{Z}_6^i + \boldsymbol{\xi}^i. \tag{5a}$$

The second model involves a hybrid of linear and nonlinear functions:

$$\boldsymbol{y}^i = 2 \cdot \mathbf{1}_T + \mathrm{Z}_1^i + 3\mathrm{Z}_2^i + 5\mathrm{Z}_3^i + 7 \cdot \sin((\mathrm{Z}_4^i)^2) + 9 \cdot \exp(-30(\mathrm{Z}_5^i - 0.5)^2) + 11\mathrm{Z}_6^i + \boldsymbol{\xi}^i, \tag{5b}$$

where the 4th and 5th components are nonlinear, and the remaining are linear.

The third model introduces a zero component into the system for comparison purposes:

$$\boldsymbol{y}^i = 2 \cdot \mathbf{1}_T + \mathrm{Z}_1^i + 3\mathrm{Z}_2^i + 5\mathrm{Z}_3^i + 7 \cdot \cos((\mathrm{Z}_4^i)^3 + \frac{\pi}{2}) + 0 \cdot \mathrm{Z}_5^i + 11\mathrm{Z}_6^i + \boldsymbol{\xi}^i, \tag{5c}$$

so that the 4th component is nonlinear, the 5th is a zero function and the remaining functions are linear. Zero functions are highlighted in this exercises as they may be harder to detect and precisely estimate and may introduce noise to our estimation.

Finally, the fourth and final model will contain one linear, one nonlinear, and four zero function components. This is the most challenging specification, given that most of the covariates are, in fact, absent from the data generating process. However, this scenario occurs in many applications with sparse datasets in various fields.

$$\boldsymbol{y}^i = 2 \cdot \mathbf{1}_T + 0 \cdot \mathrm{Z}_1^i + 3\mathrm{Z}_2^i + 0 \cdot \mathrm{Z}_3^i + 7 \cdot \cos((\mathrm{Z}_4^i)^2 + \frac{\pi}{2}) + 0 \cdot \mathrm{Z}_5^i + 0 \cdot \mathrm{Z}_6^i + \boldsymbol{\xi}^i, \tag{5d}$$

All Z variables are generated from a set of grid points starting from 0.02 with spacing of 0.02, which means the sorted values of Z are $[0.02, 0.04, 0.06, \ldots]$. All reduced-form error terms are generated from i.i.d. normal distributions with mean 0 and variance 0.2: $\xi_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.2)$. We set the number of variables $n = 3$ and the number of observations $T = 200$ for all simulation experiments.

We conduct four experiments and present the results from the first variable for each experiment. For the first experiment, all three variables are generated from equation (5a). For the second to fourth experiments, the first variable is generated from equations (5b) to (5d) respectively, while all other variables are generated from equation (5a).

First, we present the estimation results by plotting the estimated $\boldsymbol{\theta}$ values against the sorted Z values (the grid points), denoted as X^*, for each of the six components to visually assess whether our algorithm can distinguish linear components from nonlinear ones.

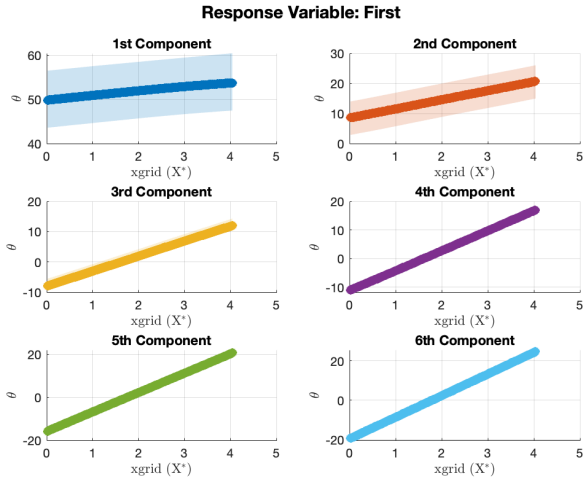

Figure 7: First response variable generated from equation (5a) where all components are linear. Plots show six components with posterior means and 90% point-wise credible interval bands.

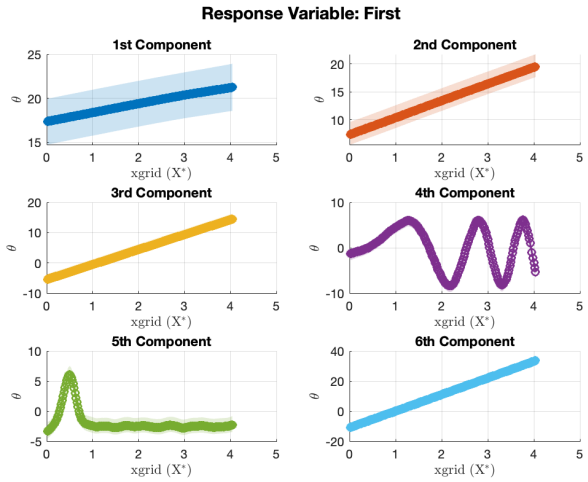

Figure 8: First response variable generated from equation (5b) where the 4th and 5th components are nonlinear. Plots show six components with posterior means and 90% point-wise credible interval bands.

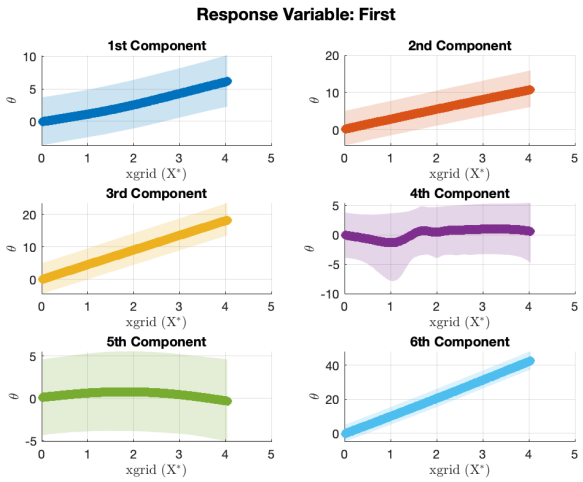

Figure 9: First response variable generated from equation (5c) where the 4th component is nonlinear and the 5th is a zero function component. Plots show six components with posterior means and 90% point-wise credible interval bands.

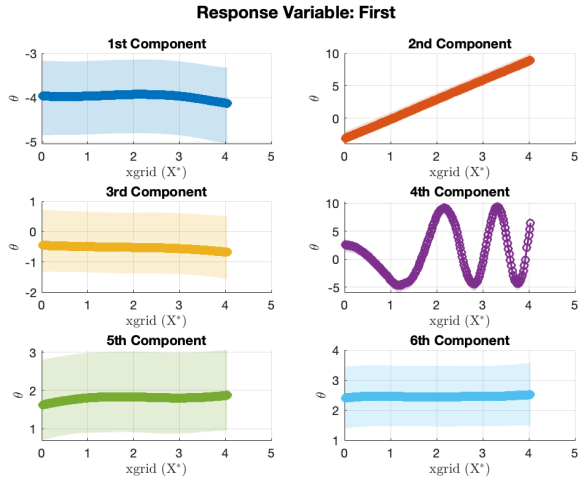

Figure 10: First response variable generated from equation (5d) where the 2nd component is linear, the 4th component is nonlinear, and the rest are zero function components. Plots show six components with posterior means and 90% point-wise credible interval bands.

As evident from the figures, although the intercepts for each component are difficult to recover, the general shape—whether the component is linear, nonlinear, or a zero function—is readily identifiable even with only 200 observations. Next, we present numerical results for the slope coefficients of these estimated components and compare them to the true values. Specifically, in Table 11 to 14, we first calculate the slope coefficients based on the estimated

$\boldsymbol{\theta}$ values for each grid point in each component, then present the posterior mean along with the 5th and 95th percentile values.

Table 11: Slope coefficient estimates for Experiment 1, where the first response variable is generated from equation (5a) with all linear components. True values reflect the linear coefficients in the data generating process.

Component	True Value	Mean	5% Percentile	95% Percentile
1	1.00	0.984	0.798	1.117
2	3.00	3.034	2.928	3.064
3	5.00	4.984	4.846	5.002
4	7.00	7.039	6.798	7.041
5	9.00	9.128	8.956	9.119
6	11.00	11.109	10.903	11.064

Table 12: Slope coefficient estimates for Experiment 2, where the first response variable is generated from equation (5b) with the 4th and 5th components being nonlinear, while the remaining components are linear.

Component	True Value[a]	Mean	5% Percentile	95% Percentile
1	1.00	0.969	0.826	1.034
2	3.00	3.032	2.968	3.071
3	5.00	4.988	4.887	4.985
4	-*	−1.016	−43.135	36.757
5	-*	0.256	−21.833	23.256
6	11.00	11.127	10.956	11.094

[a]True values represent coefficients for linear components; *Nonlinear component

Table 13: Slope coefficient estimates for Experiment 3, where the first response variable is generated from equation (5c) with the 4th component being nonlinear and the 5th component being a zero function.

Component	True Value[a]	Mean	5% Percentile	95% Percentile
1	1.00	1.557	1.148	1.811
2	3.00	2.648	2.529	2.711
3	5.00	4.559	4.480	4.564
4	-*	0.166	−1.585	4.096
5	0.00†	−0.112	−0.803	0.549
6	11.00	10.748	10.271	10.891

[a]True values represent coefficients for linear components; *Nonlinear component; †Zero function

Table 14: Slope coefficient estimates for Experiment 4, where the first response variable is generated from equation (5d) with the 2nd component being linear, the 4th component being nonlinear, and all other components being zero functions. This represents the sparse model scenario.

Component	True Value[a]	Mean	5% Percentile	95% Percentile
1	$0.00^{\dagger}$	-0.041	-0.175	0.048
2	3.00	3.029	2.919	3.072
3	$0.00^{\dagger}$	-0.057	-0.140	-0.018
4	$-^{*}$	0.985	-39.043	42.032
5	$0.00^{\dagger}$	0.064	-0.039	0.238
6	$0.00^{\dagger}$	0.027	-0.039	0.129

[a]True values represent coefficients for linear components; *Nonlinear component; †Zero function

The numerical results confirm that our algorithm effectively distinguishes between linear, nonlinear, and zero function components. Linear components are accurately estimated when isolated (Experiment 1), while nonlinear components are correctly identified through their highly variable slope estimates and wide credible intervals (Experiments 2-4). Although the presence of complex nonlinear components with rapidly changing local slopes can slightly reduce the estimation accuracy of linear components, the algorithm maintains robust performance even in sparse settings where most components are zero functions.

References

Anderson, H. M., and F. Vahid (1998): "Testing multiple equation systems for common nonlinear components," *Journal of Econometrics*, 84(1), 1–36.

Buja, A., T. Hastie, and R. Tibshirani (1989): "Linear Smoothers and Additive Models," *Ann. Statist.*, 17(2), 453–510.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2024): "Addressing COVID-19 Outliers in BVARs with Stochastic Volatility," *The Review of Economics and Statistics*, 106(5), 1403–1417.

Chan, J. C. C., G. Koop, D. J. Poirier, and J. L. Tobias (2019): *Bayesian Econometric Methods*. Cambridge University Press, 2 edn.

Chib, S., and I. Jeliazkov (2006): "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data," *Journal of the American Statistical Association*, 101(474), 685–700.

Clark, T. E., F. Huber, G. Koop, M. Marcellino, and M. Pfarrhofer (2023): "TAIL FORECASTING WITH MULTIVARIATE BAYESIAN ADDITIVE REGRESSION TREES," *International Economic Review*, 64(3), 979–1022.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3(1), 1–100.

Gefang, D., and R. Strachan (2009): "Nonlinear Impacts of International Business Cycles on the UK–A Bayesian Smooth Transition VAR Approach," *Studies in Nonlinear Dynamics & Econometrics*, 14(1), 1–33.

Giraitis, L., G. Kapetanios, and T. Yates (2014): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179(1), 46–65.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57(2), 357–384.

Hauzenberger, N., F. Huber, M. Marcellino, and N. Petz (2025): "Gaussian Process Vector Autoregressions and Macroeconomic Uncertainty," *Journal of Business & Economic Statistics*, 43(1), 27–43.

Kapetanios, G., M. Marcellino, and F. Venditti (2019): "Large time-varying parameter VARs: A nonparametric approach," *Journal of Applied Econometrics*, 34(7), 1027–1049.

Koop, G., and D. J. Poirier (2004): "Bayesian variants of some classical semiparametric regression techniques," *Journal of Econometrics*, 123(2), 259 – 282, Recent advances in Bayesian econometrics.

Koop, G., D. J. Poirier, and J. Tobias (2005): "Semiparametric Bayesian inference in multiple equation models," *Journal of Applied Econometrics*, 20(6), 723–747.

Koop, G., and J. L. Tobias (2006): "Semiparametric Bayesian inference in smooth coefficient models," *Journal of Econometrics*, 134(1), 283–315.

McCracken, M., and S. Ng (2020): "FRED-QD: A Quarterly Database for Macroeconomic Research," Working Paper 26872, National Bureau of Economic Research.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140(2), 425–449.

Peters, J., J. M. Mooij, D. Janzing, and B. Schölkopf (2014): "Causal Discovery with Continuous Additive Noise Models," *Journal of Machine Learning Research*, 15(58), 2009–2053.

Psaradakis, Z., M. O. Ravn, and M. Sola (2005): "Markov switching causality and the money–output relationship," *Journal of Applied Econometrics*, 20(5), 665–683.

Sims, C. A. (1980): "Macroeconomics and reality," *Econometrica*, 48, 1–48.

Stock, J. H., and M. W. Watson (1998): "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," Working Paper 6607, National Bureau of Economic Research.

Teräsvirta, T. (1994): "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the american Statistical association*, 89(425), 208–218.

Tibshirani, R. (2023): "Forecast Scoring and Calibration," Lecture notes for Advanced Topics in Statistical Learning, Spring 2023.

Tobias, J. L., and T. N. Bond (2026): "Semiparametric Bayesian estimation in an ordinal probit model with application to life satisfaction across countries, age and gender," *Journal of Econometrics*, 256, 105917.

Tobias, J. L., and J. C. C. Chan (2019): "An Alternate Parameterization for Bayesian Nonparametric/Semiparametric Regression," in *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part B*, vol. 40B, pp. 47–64. Emerald Publishing Ltd.

Wu, F. C. (2024): "A high-dimensional additive nonparametric model," *Journal of Economic Dynamics and Control*, 166, 104916.