Strathclyde
Discussion Papers in Economics

Bayesian Modelling of TVP-VARs Using Regression Trees*

Niko Hauzenberger,[1,2] Florian Huber,[1] Gary Koop,[2] and James Mitchell[3]

[1]University of Salzburg

[2]University of Strathclyde

[3]Federal Reserve Bank of Cleveland

No. 23 – 8

Bayesian Modeling of TVP-VARs
Using Regression Trees[*]

NIKO HAUZENBERGER,[1,2] FLORIAN HUBER,[1]
GARY KOOP,[2] and JAMES MITCHELL[3]

[1]*University of Salzburg*
[2]*University of Strathclyde*
[3]*Federal Reserve Bank of Cleveland*

Abstract. In light of widespread evidence of parameter instability in macroeconomic models, many time-varying parameter (TVP) models have been proposed. This paper proposes a nonparametric TVP-VAR model using Bayesian additive regression trees (BART) that models the TVPs as an unknown function of effect modifiers. The novelty of this model arises from the fact that the law of motion driving the parameters is treated nonparametrically. This leads to great flexibility in the nature and extent of parameter change, both in the conditional mean and in the conditional variance. Parsimony is achieved through adopting nonparametric factor structures and use of shrinkage priors. In an application to US macroeconomic data, we illustrate the use of our model in tracking both the evolving nature of the Phillips curve and how the effects of business cycle shocks on inflation measures vary nonlinearly with changes in the effect modifiers.

JEL: C11, C32, C51, E31, E32

KEYWORDS: Bayesian vector autoregression; Time-varying parameters; Nonparametric modeling; Machine learning; Regression trees; Phillips curve; Business cycle shocks

1 Introduction

Econometric models used in macroeconomics have traditionally been linear and homoskedastic. Examples include vector autoregressions (VARs), dynamic factor models (DFMs), and linearized dynamic stochastic general equilibrium (DSGE) models, inter alia. However, in recent decades, there has been growing awareness of the empirical need to allow for parameter change and to relax homoskedasticity assumptions.

Most of the models used to capture such parameter change allow for parameters to vary over time, but at a specific point in time the model remains linear. Structural break models, Markov switching models (see Sims and Zha, 2006), and time-varying parameter (TVP) models are some prominent examples. TVP regressions and TVP-VARs, in particular, have been highly successful for structural macroeconomic analysis and forecasting (see, for example, Dangl and Halling, 2012; D'Agostino et al., 2013; Koop and Korobilis, 2013; Belmonte et al., 2014; Bitto and Frühwirth-Schnatter, 2019; Korobilis, 2021; Huber et al., 2021; Hauzenberger et al., 2022). In the TVP-VAR literature, parameters are assumed to evolve according to random walks or autoregressive processes.[1] This also holds true for stochastic volatility processes, the addition of which is the most common way of relaxing the homoskedasticity assumption.

A drawback of these parametric models of parameter change is that they risk misspecification. That is, there is uncertainty about the specific law of motion driving changes in coefficients and volatilities. This not only includes how coefficients (volatilities) evolve over time but also whether changes in the law of motion depend on other observed factors (resulting in nonlinear interactions). Since policy makers often have a keen interest in how structural factors affect not only observed quantities (such as output or inflation) but also latent quantities that are often highly nonlinear functions of the reduced-form parameters of a TVP-VAR (such as the long-run unconditional mean, impulse response functions, or predictive distributions), this ignorance of possible relations between the parameters of a model and additional covariates is not merely a statistical problem but

---

[1]It is also common to assume these random walks are independent of one another. But this assumption can translate into overfitting, since coefficients often feature substantial co-movement (see Stevanovic, 2016; Chan et al., 2020).

also has substantial practical relevance. For instance, a policy maker might want to know whether the effectiveness of policy measures depend on the state of the business cycle or whether macroeconomic trends (such as trend inflation or the equilibrium interest rate) depend dynamically on other factors.

The model developed in this paper allows for flexibility in how the parameters and the error volatilities in a VAR evolve over time. The model also allows for a (possibly) nonlinear relationship between a set of covariates, labeled effect modifiers, and the reduced-form parameters of the VAR. As will be explained, this simplifies use and interpretation of the model. Our model builds on those in two recent papers, Deshpande et al. (2020) and Coulombe (2020), which also model the dynamic evolution of parameters in a nonparametric manner.[2] Both of these papers — that unlike this paper focus on univariate models — assume that each coefficient has its own nonparametric law of motion. Deshpande et al. (2020) approximate the law of motion using Bayesian additive regression trees (BART, Chipman et al., 2010), while Coulombe (2020) uses random forests. However, this flexibility might translate into overfitting and a lack of scalability to high dimensions. Given our interest in (potentially) large-scale TVP-VARs with many more parameters, we address these issues by introducing further restrictions and Bayesian shrinkage priors so as both to lessen the risk of overfitting and to maintain computational tractability.

In particular, the main contribution of the paper is the development of a flexible TVP-VAR that has several key model features that are important for inference. First, it allows for any time-variation in the parameters to be driven by a low number of latent factors. Second, it allows for factor stochastic volatility in the reduced-form VAR shocks. Third, it is nonparametric along several key dimensions: latent factors which define the law of motion of parameters (including both the TVPs and the error variance-covariance matrix) follow independent BART models. These BART models, in turn, depend on additional covariates that can be exogenous, endogenous, or latent. Novel shrinkage priors control for overfitting and allow for selection of an appropriate number of BART models.

---

[2]Fischer et al. (2023) propose a parametric alternative that controls for uncertainty with respect to the form of the state evolution equation of a TVP-VAR.

One advantage of our TVP-VAR, also noted by Coulombe (2020) in the univariate case, is that, by giving a nonparametric treatment to the parameters rather than the variables, our model remains conditionally linear in the parameters. This helps avoid the black-box nature of other nonparametric techniques, such as BART, since it facilitates interpretation of the model. Since we consider TVP-VARs and use them to carry out impulse response analysis, we can, for example, use our model to ask questions such as how impulse responses to structural shocks change if the effect modifiers are altered. This permits scenario analysis that is otherwise not possible (even with fully fledged nonparametric models) or strongly depends on the assumed relationship between the parameters and the effect modifiers.[3] Our model remains agnostic on this relationship and thus can capture nonlinearities of arbitrary form. In addition, the heteroskedastic factor structure facilitates structural identification of the VAR model (see Korobilis, 2022; Chan et al., 2022).

We also devise a new efficient and scalable Markov chain Monte Carlo (MCMC) algorithm. The key features of our algorithm are that we simulate the trees of the TVPs marginally and thus avoid mixing issues that arise if one is sampled conditionally on the other. The factor model assumed for the shocks not only ensures parsimony but also allows us to exploit computational gains by rendering the different equations of the model conditionally independent. Hence, we can use fast equation-by-equation updating.

To illustrate use of the model, we revisit the debate on the possibly evolving nature of the Phillips curve in the US. Given heteroskedasticity, we first identify the error volatility factors up to sign and scale (Chan et al., 2022), and then isolate a business cycle shock in a narrative fashion, by assuming that it is the factor that explains the largest share of variation in the reduced-form VAR residuals to output and unemployment variations during recessionary periods. Using this identified shock, we investigate the dynamic reactions of a panel of macroeconomic quantities with a particular focus on prices. Considering how different effect modifiers impact the posterior distributions of the impulse responses reveals that the effects of business cycle shocks vary nonlinearly with uncertainty and

---

[3]Particular examples are Hubrich and Tetlow (2015), Aastveit et al. (2017), Caggiano et al. (2017), Alessandri and Mumtaz (2019), and Hauzenberger et al. (2021).

according to whether the economy is in a recessionary regime.

The remainder of the paper is structured as follows. Section 2 introduces the general econometric framework. This section includes information on the likelihood function and our nonparametric treatment of the TVPs. Section 3 introduces our prior setup and discusses our posterior simulation algorithm, while Section 4 illustrates our techniques using US data. The final section concludes.

2 Nonparametric time-varying parameter VARs

In this section we develop our flexible econometric model. Our goal is to model the evolution of an $M \times 1$-vector of macroeconomic time series, which we denote by $\{\boldsymbol{y}_t\}_{t=1}^T$. The elements in $\boldsymbol{y}_t$ might feature structural breaks, changing cross-variable dependencies, and/or different persistence behavior over time. In addition, it could be that the number of time series M is large. Our framework will be capable of simultaneously handling many time series that might feature complex dynamics.

2.1 A standard time-varying parameter model

We assume that $\boldsymbol{y}_t$ evolves according to a VAR model with drifting parameters. Our TVP-VAR with P lags is given by:

$$\boldsymbol{y}_t = \sum_{p=1}^{P} \left(\boldsymbol{A}_p + \boldsymbol{B}_{pt}\right) \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{\Omega}_t). \tag{1}$$

Note that we have written the TVP-VAR as involving $\boldsymbol{A}_p$ for $p = 1, .., P$, which are $M \times M$-matrices of constant coefficients, and $\boldsymbol{B}_{pt}$ for $p = 1, .., P$, which are $M \times M$-matrices of TVPs capturing deviations from the constant part of the model. $\boldsymbol{\epsilon}_t$ is an $M \times 1$-vector of Gaussian shocks with mean zero and time-varying variance-covariance matrix $\boldsymbol{\Omega}_t$.

There is substantial evidence that macroeconomic data are driven by a small set of fundamental shocks (Bai and Ng, 2007). We incorporate this insight into our model by

assuming that $\boldsymbol{\epsilon}_t$ features a common factor structure:

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Gamma}\boldsymbol{q}_t + \boldsymbol{\varepsilon}_t \quad \Leftrightarrow \quad \boldsymbol{\Omega}_t = \boldsymbol{\Gamma}\boldsymbol{R}_t\boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_t, \tag{2}$$

where $\boldsymbol{q}_t$ is a $Q_q(\ll M)\times 1$-vector of Gaussian-distributed factors with mean zero and diagonal time-varying variance-covariance matrix $\boldsymbol{R}_t = \text{diag}(r_{1t}, \ldots, r_{Q_qt})$. $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{Q_q})$ refers to a $M \times Q_q$-matrix of factor loadings, while $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Mt})'$ is an $M \times 1$-vector of Gaussian idiosyncratic shocks with mean zero and diagonal variance-covariance matrix $\boldsymbol{\Sigma}_t = \text{diag}(\sigma_{1t}^2, \ldots, \sigma_{Mt}^2)$. One key observation is that, conditional on the factors $\boldsymbol{q}_t$, the shocks $\boldsymbol{\varepsilon}_t$ are independent and equation-by-equation estimation is possible. This factor structure has been used in other papers to facilitate estimation of large VARs (Kastner and Huber, 2020; Chan, Forthcoming; Clark et al., Forthcoming) and, in addition, this structure can also facilitate identification of the factors as structural VAR disturbances (Korobilis, 2022; Chan et al., 2022). Moreover, in contrast to VAR-based estimation using a Cholesky decomposition of the error covariances, another convenient feature of our model is that it is invariant to how the variables are ordered in $\boldsymbol{y}_t$ (for a formal argument, see, for example, Chan et al., 2022).

It is worth stressing that the factor model, without additional restrictions, is not point identified. This is because the factors and the loadings enter the likelihood in product form and are, thus, not invariant to rotation, column and sign switching. Point identification can then be achieved by introducing restrictions on the loadings. A standard restriction assumes that the first $Q_q \times Q_q$ leading matrix of $\boldsymbol{\Gamma}$ is lower uni-triangular. This immediately implies that the resulting estimates will depend on the ordering of the elements in $\boldsymbol{y}_t$, a property that we would like to avoid. Hence, in what follows we do not impose identification restrictions on our factor model during MCMC estimation. Results in Chan et al. (2022) suggest that the decomposition in Eq. (2) is identified up to column and sign switching. Since our focus is on impulse responses to changes in particular elements in $\boldsymbol{q}_t$, we tackle column and sign switching ex-post. Kaufmann and Schumacher (2019) follow a similar strategy, post-processing the posterior factor draws to point-identify the factors/shocks. We discuss this detail further in the empirical

application below.

Up to this point we have remained silent on how the latent states (which include both the VAR coefficients and the time-varying elements of the error variances) evolve over time. In the next section, we introduce a flexible law of motion for the latent states.

2.2   A nonparametric law of motion for the TVPs

A standard assumption in the TVP-VAR literature is that the elements in $\boldsymbol{B}_{pt}$ ($p = 1, \ldots, P$) and $\boldsymbol{R}_t$ evolve according to simple parametric stochastic processes, most often random walks (Primiceri, 2005; Cogley and Sargent, 2005; Belmonte et al., 2014; Bitto and Frühwirth-Schnatter, 2019). Assuming that the states evolve according to random walks introduces parsimony, because it implies a prior on the smoothness of the time variation in the coefficients. However, in turbulent periods, such as during the global financial crisis or the COVID-19 pandemic, it could be that parameters change rapidly and display sharp structural breaks. In such a case, a mixture model that models the evolution of the parameters as characterized by a low number of breaks (Sims and Zha, 2006; Koop and Potter, 2007; Kaufmann, 2015) would be more appropriate. Another possibility is that parameter change could depend on exogenous *effect modifiers*, translating into a specification with interaction effects. The nonparametric approach that we develop allows for all these possibilities.

Although most TVP-VARs allow for each coefficient to have its own random walk process, this is probably too flexible. That is, it is an empirical regularity that there is a high degree of co-movement in the parameters. This motivates the inclusion of a factor structure in the TVPs (that is, allowing for the the process innovation variance-covariance matrix to be of reduced-rank). In the parametric TVP-VAR literature, Chan et al. (2020) propose a model that assumes a factor structure on the TVPs and assumes that the factors driving the states evolve according to a random walk. Fischer et al. (2023) modify this approach by allowing for different forms of parameter change. This is achieved through including effect modifiers that can be either observed or latent.

In this paper, we do something similar, but we do it nonparametrically. That is, we

assume there are a small number of latent nonlinear factors driving parameter change, which we estimate nonparametrically. In other words, we remain agnostic on the precise law of motion of the latent states, and let the data decide on the appropriate state dynamics, while achieving parsimony by introducing a factor structure to the TVPs.

We begin by writing the TVP-VAR in more compact form. Let $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \dots, \boldsymbol{y}'_{t-p})'$ denote a $K = (MP) \times 1$-vector of covariates. Moreover, let $\boldsymbol{A} = (\boldsymbol{A}_1, \dots, \boldsymbol{A}_P)$ and $\boldsymbol{B}_t = (\boldsymbol{B}_{1t}, \dots, \boldsymbol{B}_{Pt})$ refer to $M \times K$-matrices that stack the VAR coefficients. Since our model, conditional on the latent factors, is a system of independent regression models we can focus on the mth equation of $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{x}_t$. This regression model can be expressed as:

$$\tilde{y}_{mt} = \boldsymbol{x}'_t \boldsymbol{\beta}_{mt} + \boldsymbol{q}'_t \boldsymbol{\gamma}_m + \varepsilon_{mt}, \quad \boldsymbol{q}_t \sim \mathcal{N}(\boldsymbol{0}_{Q_q}, R(\boldsymbol{z}_t)), \quad \varepsilon_{mt} \sim \mathcal{N}(0, \sigma^2_{mt}). \tag{3}$$

Here, $\boldsymbol{\beta}_{mt}$ and $\boldsymbol{\gamma}_m$ refer to the mth rows of $\boldsymbol{B}_t$ and $\boldsymbol{\Gamma}$, respectively. The factors $\boldsymbol{q}_t$ arise from a Gaussian distribution with variance $\boldsymbol{R}_t = R(\boldsymbol{z}_t) = \text{diag}(r_1(\boldsymbol{z}_t), \dots, r_{Q_q}(\boldsymbol{z}_t))$ where $r_s : \mathbb{R}^N \to \mathbb{R}^+$ is an unknown function. Notice that the error variances depend on a set of effect modifiers in $\boldsymbol{z}_t$.

We assume that $\boldsymbol{\beta}_{mt}$ evolves according to:

$$\boldsymbol{\beta}_{mt} = \boldsymbol{\Lambda}_m F_m(\boldsymbol{z}_t) + \boldsymbol{\eta}_{mt}, \tag{4}$$

where $F_m(\boldsymbol{z}_t) = (f_{m1}(\boldsymbol{z}_t), \dots, f_{mQ_\beta}(\boldsymbol{z}_t))'$ denotes an unknown function with Q_β components $f_{mq} : \mathbb{R}^N \to \mathbb{R}$, $\boldsymbol{\Lambda}_m$ is a $K \times Q_\beta$-matrix of factor loadings, and Q_β is the number of latent factors that drive the TVPs. In addition, we assume that $\boldsymbol{\eta}_{mt} \sim \mathcal{N}(\boldsymbol{0}_K, \boldsymbol{V}_m)$ is a vector of Gaussian shocks with $\boldsymbol{V}_m = \text{diag}(v^2_{m1}, \dots, v^2_{mK})$ denoting the process innovation variances.

Conditional on choosing an appropriate number of factors Q_β, this specification is extremely flexible. It allows for (potentially) nonlinear interactions between $\boldsymbol{z}_t$ and $\boldsymbol{\beta}_{mt}$ (and thus implicitly $\boldsymbol{x}_t$). If elements in $\boldsymbol{\beta}_{mt}$ do not depend on $\boldsymbol{z}_t$ the corresponding loadings are zero and time-variation can still be captured through the presence of the

idiosyncratic shocks in $\boldsymbol{\eta}_{mt}$.[4] Notice that the variances in $\boldsymbol{V}_m$ also control the weight put on the nonlinear factor component. For instance, if the jth coefficient closely co-moves with the other coefficients in a nonlinear manner, v_{mj}^2 will be close to zero.

To make this model operational we have to learn the functions F and R and decide on appropriate effect modifiers in $\boldsymbol{z}_t$. Our approach remains agnostic on the specific shape of both F and R and uses BART to estimate them. The next sub-sections show how this is achieved.

The choice of effect modifiers should depend on the application. The modifiers could include exogenous regressors, deterministic functions of time, lagged elements of $\boldsymbol{y}_t$, or latent quantities. If elements in $\boldsymbol{z}_t$ are endogenous and interest centers on higher-order impulse responses or multi-step-ahead predictive densities, one could either set up a separate law of motion for $\boldsymbol{z}_t$ or introduce hard restrictions on how the $\boldsymbol{z}_t's$ are expected to evolve over the forecast/impulse response horizon. In our empirical application below, we follow the latter approach, not only for simplicity, but because we are interested in how the dynamic reactions of $\boldsymbol{y}_t$ to shocks depend on the elements in $\boldsymbol{z}_t$ taking on certain values.[5] This enables us to answer what-if questions, such as, "How would inflation react to business cycle shocks if uncertainty is (and remains) high?", or, "How do price reactions to business cycle movements change if the population becomes increasingly over-aged?"

Another interesting possibility would be to set $\boldsymbol{z}_t = \boldsymbol{x}_t$. In this case, however, interpretation becomes more difficult since the model then becomes nonlinear in $\boldsymbol{y}_t$. This would then necessitate the use of generalized impulse responses (Koop et al., 1996) to carry out dynamic analysis. With our application seeking to characterize features of the US business cycle, we choose effect modifiers that either slowly evolve independently of the business cycle, like factors related to the age-composition of the population, or binary variables (such as recession indicators), or other variables not included in $\boldsymbol{y}_t$. One such variable we consider is the uncertainty measure proposed in Jurado et al. (2015) which can be interpreted as a proxy of (unobserved) macroeconomic uncertainty.

---

[4]Our model can also be related to random coefficient models, see Frühwirth-Schnatter et al. (2004).
[5]This would resemble common practice in, e.g., threshold or Markov switching models that condition on the prevailing regime when computing impulse responses.

Another strategy to selecting the elements of $\boldsymbol{z}_t$ would be to entertain a large set of potential effect modifiers, and then use regularization techniques. As we will describe below, our approach is capable of handling all these cases without additional modification.

2.3 Learning the unknown functions using BART

We approximate each function f_{mq} through a sum-of-trees model (Chipman et al., 2010):

$$f_{mq} \approx g(\boldsymbol{Z}|\mathcal{T}^{\beta}_{mq}, \boldsymbol{\mu}_{mq}) = \sum_{s=1}^{S_{\beta}} u\left(\boldsymbol{Z}|\mathcal{T}^{\beta}_{mq,s}, \boldsymbol{\mu}_{mq,s}\right), \tag{5}$$

with the $T \times N$-matrix $\boldsymbol{Z}$ having a typical t^{th} row $\boldsymbol{z}'_t$ and u being a regression tree function that depends on a tree structure, $\mathcal{T}^{\beta}_{mq} = \{\mathcal{T}^{\beta}_{mq,1}, \ldots, \mathcal{T}^{\beta}_{mq,S_{\beta}}\}$, which is a sequence of disjoint sets that partition the input space and a vector of terminal node parameters $\boldsymbol{\mu}_{mq} = \{\boldsymbol{\mu}_{mq,1}, \ldots, \boldsymbol{\mu}_{mq,S_{\beta}}\}$ of dimension b_{mq}. These partitions are driven by splitting rules of the form $z_{jt} \leq c_j$ or $z_{jt} > c_j$, with z_{jt} denoting the jth element of $\boldsymbol{z}_t$ and c_j being a threshold parameter. Moreover, S_{β} is the number of trees used to approximate each of the functions (factors) f_{mq}. Eq. (5) is a standard BART model. To avoid issues associated with overfitting when S_{β} is large, Chipman et al. (2010) propose using a regularization prior to force the trees to take a particularly simple form and thus explain only a small fraction of the variation of the response variable. Adding together many simple trees (weak learners) has been found to work better than working with a single more complicated tree. We follow such an approach in this paper.

Plugging (5) into (4) yields our state equation:

$$\boldsymbol{\beta}_{mt} = \sum_{q=1}^{Q_{\beta}} \boldsymbol{\lambda}_{mq} g(\boldsymbol{z}_t|\mathcal{T}^{\beta}_{mq}, \boldsymbol{\mu}_{mq}) + \boldsymbol{\eta}_{mt}, \tag{6}$$

with $\boldsymbol{\lambda}_{mq}$ denoting the qth column of $\boldsymbol{\Lambda}_m$. This shows that we combine Q_{β} BART models (each used to approximate one of the Q_{β} functions). Setting $Q_{\beta} = 1$ implies that all coefficients are driven by a single factor (if $\boldsymbol{\lambda}_{mq} \neq \boldsymbol{0}_K$), while when setting $Q_{\beta} = K$

we obtain a model closely related to the one proposed in Deshpande et al. (2020) and Coulombe (2020). Since the latter specification, in light of large K, does not scale well to high dimensions we will focus on the case where $K \gg Q_\beta$, which frequently arises in the analysis of large TVP-VAR models. We will call models that assume this nonparametric factor form for the conditional mean, TVP-FBART. Ahead of our main empirical application and to help the reader further understand our model, Sub-section C.1 in the Online Appendix provides a toy empirical example to illustrate how BART can be used to approximate TVPs.

2.4 Flexible heteroskedasticity specifications

Recall that our model also assumes that the shocks $\boldsymbol{\epsilon}_t$ feature a factor structure. We will again approximate the factor-specific functions, $r_s(\boldsymbol{z}_t)$, in $\boldsymbol{R}_t$ with BART. More precisely, our approach can be interpreted as a variant of heteroskedastic BART (heteroBART, see Pratola et al., 2020). heteroBART is a multiplicative version of BART and assumes that the trees enter the model in product form. In this paper, we follow Clark et al. (Forthcoming) and linearize the model so that standard BART techniques can be used.

Let the sth element of $\boldsymbol{q}_t$ be given by:

$$q_{st} = \exp\left(\sum_{d=1}^{S_q} u(\boldsymbol{z}_t | \mathcal{T}_{sd}^q, \boldsymbol{\pi}_{sd})/2\right) \times \xi_{st}, \quad \xi_{st} \sim \mathcal{N}(0, 1), \tag{7}$$

with S_q being the number of trees used to approximate the variance functions, where $\mathcal{T}_{sd}^q$ and $\boldsymbol{\pi}_{sd}$ denote the corresponding tree structures and terminal node parameters, respectively. To render Eq. (7) linear we square it and take logs. This yields a linear equation with shocks that are log-χ^2 distributed with one degree of freedom, a distribution which can be well approximated using a ten-component mixture approximation (see Omori et al., 2007):

$$\tilde{q}_{st} = \log(q_{st}^2) = \sum_{d=1}^{S_q} u(\boldsymbol{z}_t | \mathcal{T}_{sd}^q, \boldsymbol{\pi}_{sd}) + \tilde{\xi}_{st}, \quad \tilde{\xi}_{st} = \log(\xi_{st}^2) \sim \sum_{i=1}^{10} w_i \mathcal{N}(m_i, \mathfrak{r}_i^2). \tag{8}$$

Here, m_i, w_i, and $\mathfrak{r}_i^2$ are fixed numbers defining the mixture components taken from Table 1 in Omori et al. (2007).

This specification of heteroBART implies that the factor volatilities are allowed to change rapidly, but can also move more gradually. This feature might pay off during recessions, where large jumps in error volatilities are common. Traditional stochastic volatility models will be unable to match this pattern, since they assume that the log-volatilities evolve according to a stochastic process that translates into a more gradual evolution of the error variances. Such behavior is warranted if the trend movement in volatility is persistent (such as during the Great Moderation). For heteroBART, matching slowly evolving trends is also possible but considerably harder. To allow for smoothly evolving stochastic trends we combine heteroBART with a standard stochastic volatility model in the measurement errors (that is, the elements in $\boldsymbol{\Sigma}_t$). The combination between a parametric law of motion for $\log(\sigma_{mt}^2)$ and q_{st} allows for rich dynamics in terms of $\boldsymbol{\Omega}_t$.[6] We will use the abbreviation FHB (using a factor structure involving heteroBART) for models which adopt this specification. Thus, our most general model is TVP-FBART-FHB.

2.5 Summary of key model features

The model described in the previous sub-sections is very flexible and nests a wide variety of competing models. In this sub-section, we first summarize key model features and then discuss how our model is related to alternative models commonly used in the literature.

Flexible machine learning techniques such as BART have the shortcoming that interpretability is difficult. As noted, for example in Coulombe (2020), using regression trees to model the parameters of a TVP regression allows for flexibility but also maintains simplicity of interpretation. In our case, once we have learned the TVPs and the functions driving them using BART, interpretation of the model works analogously to a standard TVP-VAR model. Hence, one can compute functions of the parameters such as

---

[6]Another option to capture smoothly varying trends with heteroBART would be through the specification of a latent component which enters $\boldsymbol{z}_t$. But this would require nonlinear filtering algorithms or linear approximations (which can fail in certain environments) such as the ones proposed in Huber et al. (2023).

impulse responses, forecast error variance or historical decompositions, and conditional forecasts using standard techniques. This constitutes a big advantage of our approach relative to models such as the one proposed in Huber and Rossini (2022). Computation of (generalized) impulse response function in traditional BART-based VAR models is much more involved, as the model remains nonlinear.[7]

The previous paragraph is related to the effect that shocks might have on $\boldsymbol{y}_t$. Since the effect modifiers influence $\boldsymbol{y}_t$ indirectly through the BART modeling of the TVPs, we can also assess how $\boldsymbol{z}_t$ affects $\boldsymbol{y}_t$. This can be easily achieved in our framework since one can compute different realizations of the TVPs for different configurations of $\boldsymbol{z}_t$. Doing so allows us to study how (higher-order) interaction effects, which might take an unknown form, impact quantities such as forecast distributions, impulse responses, or even long-run trends such as the (time-varying) unconditional mean of the TVP-VAR. We will illustrate these features in our empirical work that follows in Section 4 below.

Apart from the ease of interpretation and the additional inferential possibilities, our model, for appropriately chosen values of Q_β, Q_q, S_β, and S_q, provides a great deal of flexibility when it comes to capturing different forms of parameter change. While our aim is to introduce as few restrictions on the state evolution as possible, we can nevertheless control the dynamics of the TVPs by choosing appropriate values of S_β and S_q. In principle, larger values of S_β and S_q are consistent with smooth law of motions of the parameters, whereas smaller values imply parameter dynamics closer to the ones generated by a structural break model. An extreme case of our model would set $Q_\beta = S_\beta = 1$. This specification would imply that parameters follow a single regression tree and are proportional to each other. In our empirical work we will explore the sensitivity of results by varying these parameters.

---

[7]Koop et al. (1996) discuss how to compute generalized impulse response functions in nonlinear multivariate models.

3 Bayesian inference

3.1 The prior

We start our discussion with the priors relating to the regression trees and the process innovation variances. The remaining priors are relatively standard and a discussion can be found in Section A in the Online Appendix.

Chipman et al. (1998) and Chipman et al. (2010) specify a tree-generating stochastic process on the tree structures, $\mathcal{T}^{\beta}_{mq,s}$ and $\mathcal{T}^{q}_{sd}$. Our approach is similar, but specifies the prior such that the probability of growing more complex trees decreases with the number of factors $\nu = 1, \ldots, Q_j$ for $j \in \{\beta, q\}$. This process is designed to penalize complex trees and consists of three features:

1. A decreasing probability that a node is non-terminal. Let $\mathfrak{n} = 0, 1, \ldots$ denote a particular node at depth $\mathfrak{n}$. We model the probability that this node is non-terminal as follows:

$$\alpha^{\nu}(1 + \mathfrak{n})^{-\zeta^{\nu}},$$

where α is between 0 and 1 and $\zeta > 1$. This implies that α acts as a base parameter and ζ penalizes more complex trees by shrinking the probability that a given node is non-terminal for higher-order nodes. A typical choice that works well for many datasets is $\alpha = 0.95$ and $\zeta = 2$ (see Chipman et al., 2010). Notice that the base probability decreases in ν, implying that for large values of ν the probability of forking a new branch of a tree decreases substantially. This effect is complemented by the shrinkage parameter ζ^{ν} which grows rapidly in ν. By setting $\nu = q$, this prior effectively allows the modeler to select Q_{β} by forcing the functions f_{mq}, for large q, towards a constant function (implying no effect on coefficient dynamics). This prior is used for $\mathcal{T}^{\beta}_{mq,s}$. For $\mathcal{T}^{q}_{sd}$, we set $\nu = 1$ (regardless of s) and thus use the benchmark prior of Chipman et al. (2010).

2. A prior distribution on the splitting variables in $\boldsymbol{z}_t$. In the absence of strong prior

14

information we follow much of the literature and use a discrete uniform prior on the elements in $\boldsymbol{z}_t$. Hence, at every node, every variable in $\boldsymbol{z}_t$ is equally likely to be used to split up the input space.

3. A prior on the thresholds c_j within a given splitting rule is assumed to be uniformly distributed. Similar to the prior on the splitting variables, this specification remains agnostic on the precise values that the thresholds may take.

This prior encourages smaller trees and is thus consistent with the notion that each individual tree is a "weak learner," but the composite model is capable of capturing complex dynamics in the parameters.

The prior on the terminal node parameters is Gaussian. Following Chipman et al. (2010), we scale the data such that the dependent variable is between -0.5 and 0.5 and our prior covers this range. Let $\mu_{mq,ij}$ denote the jth element of $\boldsymbol{\mu}_{mq,i}$ and $\pi_{sd,j}$ the jth element of $\boldsymbol{\pi}_{sd}$. The prior for the respective element is then given by:

$$\mu_{mq,ij} \sim \mathcal{N}\left(0, \frac{1}{2\kappa S_\beta}\right) \quad \text{and} \quad \pi_{sd,j} \sim \mathcal{N}\left(0, \frac{1}{2\kappa S_q}\right).$$

Here, κ is a parameter that controls the prior variance. Shrinkage is introduced by increasingly forcing $\mu_{mq,ij}$ ($\pi_{sd,j}$) towards zero if S_β (S_q) is large. Since S_j ($j \in \{\beta, q\}$) is typically between 50 and 200, this prior is the second ingredient of BART used to capture the notion that each tree explains only a small amount of variation in $\boldsymbol{\beta}_t$ (and q_{st}).

On the different elements of $\boldsymbol{V}_m$, several priors are possible. The simple conjugate inverse Gamma prior can be used. This prior, however, has implications for our model, since it rules out values of v_{mj}^2 very close to zero. Hence, it would artificially push the likelihood away from the factor part in Eq. (4). We follow recommendations in Frühwirth-Schnatter and Wagner (2010) and use a prior that introduces shrinkage on $\boldsymbol{V}_m$. Our prior assumes that v_{mj}^2 arises from a Gamma distribution:

$$v_{mj}^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_v}\right) \quad \Leftrightarrow \quad \pm v_{mj} \sim \mathcal{N}(0, B_v),$$

with B_v being a scalar hyperparameter that controls the amount of shrinkage towards a factor structure in the TVPs. Since there exists strong evidence that the TVPs feature a factor structure, we set $B_v = 0.01$ to have a tight prior on the idiosyncratic deviations of the TVPs from the common factor structure.

3.2 Markov chain Monte Carlo sampling

We sample from the joint posterior distribution of the model by using an MCMC algorithm that, conditional on the latent factors, simulates the coefficients and latent states for each equation separately. Since for some of the steps in the sampler we integrate out other parameters the precise ordering of the steps of the MCMC algorithm is important to simulate from the correct stationary distribution. Our algorithm cycles between the following steps. For each equation $m = 1, \ldots, M$:

1. **Sampling the trees.** We sample the regression trees associated with the VAR coefficients marginally of the TVPs and conditional on the remaining parameters and latent states. Similar to Chipman et al. (2010), the trees are simulated on a tree-by-tree basis. Specifically, to sample the qth tree conditional on the other trees we first integrate out the TVPs of the mth equation by plugging the state equation in Eq. (6) into the observation equation in Eq. (3). Subtracting $\boldsymbol{q}_t' \boldsymbol{\gamma}_m$ from $\tilde{y}_{mt}$ and carrying out some algebraic manipulations yields:

$$\frac{\tilde{y}_{mt}^* - \sum_{j \neq q} \tilde{x}_{mj,t} g(\boldsymbol{z}_t | \mathcal{T}_{mj}^\beta, \boldsymbol{\mu}_{mj})}{\tilde{x}_{mq,t}} = g(\boldsymbol{z}_t | \mathcal{T}_{mq}^\beta, \boldsymbol{\mu}_{mq}) + \tilde{\varsigma}_{mt},$$

with $\tilde{y}_{mt}^* = \tilde{y}_{mt} - \boldsymbol{q}_t' \boldsymbol{\gamma}_m$, $\tilde{\boldsymbol{x}}_{mt} = \boldsymbol{x}_t' \boldsymbol{\Lambda}_m$ and $\tilde{x}_{mq,t}$ denoting the qth element of $\tilde{\boldsymbol{x}}_{mt}$. Moreover, $\tilde{\varsigma}_{mt} \sim \mathcal{N}\left(0, \frac{\boldsymbol{x}_t' \boldsymbol{V}_m \boldsymbol{x}_t + \sigma_{mt}^2}{\tilde{x}_{mq,t}^2}\right)$ refers to a period-specific (independent) white noise shock. This is a nonparametric regression model and the Metropolis Hastings algorithm proposed in Chipman et al. (1998) can be used to simulate the tree structures.

2. **Sampling the terminal node parameters.** Conditional on the tree structures we can obtain the terminal node parameters by sampling from univariate Gaussian posterior distributions. The corresponding moments take particularly simple forms

since, conditional on a tree structure that allocates observations to a specific terminal node, the posterior resembles the one of a simple intercept model under a conjugate Gaussian prior.

3. **Sampling the coefficient factor loadings.** We sample the factor loadings in $\mathbf{\Lambda}_m$ conditional on the estimated trees, the process innovation variances in $\mathbf{V}_m$, the latent factors $\mathbf{q}_t$ but marginally of the TVPs. This can be achieved as follows. The observation equation, after integrating out the TVPs, can be written as:

$$\tilde{y}_{mt}^* = \mathbf{x}_t' \mathbf{\Lambda}_m F_m(\mathbf{z}_t) + \varsigma_{mt}, \quad \varsigma_{mt} \sim \mathcal{N}\left(0, \mathbf{x}_t' \mathbf{V}_m \mathbf{x}_t + \sigma_{mt}^2\right).$$

Notice that $\mathbf{x}_t' \mathbf{\Lambda}_m F_m(\mathbf{z}_t)$ equals $(F_m'(\mathbf{z}_t) \otimes \mathbf{x}_t') \mathrm{vec}(\mathbf{\Lambda}_m)$, implying a standard multivariate regression model. Hence, under the Horseshoe prior, $\mathrm{vec}(\mathbf{\Lambda}_m)$ follows a multivariate Gaussian posterior distribution with posterior mean and variance taking standard forms.

4. **Sampling the TVPs.** To sample the full history of $\{\boldsymbol{\beta}_{mt}\}_{t=1}^T$, we exploit the static representation of the model in Eq. (3). Let $\tilde{\mathbf{y}}_m^\bullet$ denote a $T \times 1$-vector with typical element $\tilde{y}_{mt}^\bullet = (\tilde{y}_{mt} - \mathbf{q}_t' \boldsymbol{\gamma}_m)/\sigma_{mt}$ and $\mathbf{W}_m = \mathrm{bdiag}(\mathbf{x}_1'/\sigma_{m1}, \ldots, \mathbf{x}_T'/\sigma_{mT})$ is a $T \times TK$-dimensional block diagonal matrix with the normalized $\mathbf{x}_t'$'s along its main diagonal. The corresponding static representation of the model is:

$$\tilde{\mathbf{y}}_m^\bullet = \mathbf{W}_m \boldsymbol{\beta}_m + \tilde{\boldsymbol{\varepsilon}}_m, \quad \tilde{\boldsymbol{\varepsilon}}_m \sim \mathcal{N}(\mathbf{0}_T, \mathbf{I}_T),$$

with $\boldsymbol{\beta}_m = (\boldsymbol{\beta}_{m1}', \ldots, \boldsymbol{\beta}_{mT}')'$ and $\tilde{\boldsymbol{\varepsilon}}_m = (\varepsilon_{m1}, \ldots, \varepsilon_{mT})'$. The corresponding posterior distribution of $\boldsymbol{\beta}_m$ is a TK-dimensional Gaussian distribution with a block-diagonal posterior covariance matrix:

$$\boldsymbol{\beta}_m | \bullet \sim \mathcal{N}(\overline{\boldsymbol{\beta}}_m, \overline{\mathbf{V}}_{\boldsymbol{\beta}_m}), \quad \text{where}$$

$$\overline{\mathbf{V}}_{\boldsymbol{\beta}_m} = \left(\mathbf{W}_m' \mathbf{W}_m + (\mathbf{I}_T \otimes \mathbf{V}_m^{-1})\right)^{-1} \quad \text{and} \quad \overline{\boldsymbol{\beta}}_m = \overline{\mathbf{V}}_{\boldsymbol{\beta}_m} \left(\mathbf{W}_m' \tilde{\mathbf{y}}_m^\bullet + (\mathbf{I}_T \otimes \mathbf{V}_m^{-1}) \underline{\boldsymbol{\beta}}_m\right).$$

The TK-dimensional prior mean vector is given by $\underline{\boldsymbol{\beta}}_m = ((\mathbf{\Lambda}_m F_m(\mathbf{z}_1))', \ldots, (\mathbf{\Lambda}_m F_m(\mathbf{z}_T))')'$.

Hauzenberger et al. (2022) provide fast algorithms to sample from this posterior distribution which exploit the fact that the rank of $\boldsymbol{W}_m'\boldsymbol{W}_m$ is T.

5. **Sampling the process innovation variances.** The posterior distribution of the process innovation variances v_{mj}^2 follows a generalized inverse Gaussian (GIG) distribution:

$$v_{mj}^2|\bullet \sim \mathcal{GIG}\left(1/2 - T/2, \sum_{t=1}^{T}\eta_{mj,t}^2, (2B_v)^{-1}\right).$$

Here, $\eta_{mj,t}$ denotes the jth element of $\boldsymbol{\eta}_{mt}$.

6. **Sampling the time-invariant regression coefficients.** Conditional on the TVPs, time-varying error variances, latent factors and loadings, the time-invariant regression coefficients can be obtained from standard multivariate Gaussian posterior distributions. More precisely, the mth row of $\boldsymbol{A}$, $\boldsymbol{a}_m$, arises from a K-dimensional Gaussian posterior:

$$\boldsymbol{a}_m|\bullet \sim \mathcal{N}(\overline{\boldsymbol{a}}_m, \overline{\boldsymbol{V}}_{a_m}), \quad \text{with moments}$$

$$\overline{\boldsymbol{V}}_{a_m} = (\boldsymbol{X}'\boldsymbol{\Sigma}_m^{-1}\boldsymbol{X} + \underline{\boldsymbol{V}}_{a_m}^{-1})^{-1} \quad \text{and} \quad \overline{\boldsymbol{a}}_m = \overline{\boldsymbol{V}}_{a_m}\boldsymbol{X}'\boldsymbol{\Sigma}_m^{-1}\overline{\boldsymbol{y}}_m.$$

Here, $\underline{\boldsymbol{V}}_{a_m}$ is the diagonal prior covariance matrix with typical element given by the Horseshoe prior described in the previous sub-section and $\overline{\boldsymbol{y}}_m$ is a $T \times 1$-vector with t^{th} element $(y_{mt} - \boldsymbol{x}_t'\boldsymbol{\beta}_{mt} - \boldsymbol{q}_t'\boldsymbol{\gamma}_m)$.

7. **Sampling the factor loadings $\boldsymbol{\gamma}_m$.** The factor loadings $\boldsymbol{\gamma}_m$ can be straightforwardly obtained by estimating a regression model with covariates $\boldsymbol{q}_t$, response variable $\tilde{y}_{mt} - \boldsymbol{x}_t'\boldsymbol{\beta}_{mt}$, and heteroskedastic shocks with variances $\boldsymbol{\Sigma}_m$. The corresponding posterior distribution of $\boldsymbol{\gamma}_m$ is Gaussian and the moments take a form similar to the ones in Step 6.

8. **Sampling the hyperparameters associated with the prior on $\boldsymbol{a}_m = (a_{1m}, \ldots, a_{Km})'$.** To sample the diagonal elements of $[\underline{\boldsymbol{V}}_{a_m}]_{ii} = \varrho_{i,a_m}^2\varpi_{a_m}^2$, with ϱ_{i,a_m} and ϖ_{a_m} denoting the ith local shrinkage parameter and the equation-specific global shrinkage param-

eter, we use the efficient and simple-to-implement sampler proposed in Makalic and Schmidt (2015). This sampler introduces two types of auxiliary random variables, $r_{i,a_m} \sim \mathcal{G}^{-1}(1/2, 1)$ and $n_{a_m} \sim \mathcal{G}^{-1}(1/2, 1)$, that have inverse Gamma priors. Simulating from the posterior of ϱ^2_{i,a_m} and $\varpi^2_{a_m}$ is then achieved by first simulating r_{i,a_m} and n_{a_m} from inverse Gamma distributions:

$$r_{i,a_m}|\bullet \sim \mathcal{G}^{-1}\left(1, 1 + \frac{1}{\varrho^2_{i,a_m}}\right) \quad \text{and} \quad n_{a_m}|\bullet \sim \mathcal{G}^{-1}\left(1, 1 + \frac{1}{\varpi^2_{a_m}}\right).$$

Conditionally on a draw of r_{i,a_m}, n_{a_m}, the full conditional posterior of ϱ^2_{i,a_m} and $\varpi^2_{a_m}$ is inverse Gamma as well:

$$\varrho^2_{i,a_m}|\bullet \sim \mathcal{G}^{-1}\left(1, \frac{1}{r_{i,a_m}} + \frac{a^2_{im}}{2\varpi^2_{a_m}}\right) \quad \text{and} \quad \varpi^2_{a_m}|\bullet \sim \mathcal{G}^{-1}\left(1, \frac{1}{n_{a_m}} + \frac{1}{2}\sum_{i=1}^{K} \frac{a^2_{im}}{\varrho^2_{i,a_m}}\right).$$

9. **Sampling the hyperparameters associated with the prior on $\boldsymbol{\Lambda}_m$.** This step closely mirrors Step 7, with the relevant quantities being replaced by the corresponding elements in $\boldsymbol{\Lambda}_m$.

10. **Sampling the latent log-volatilities and the parameters of the state equation.** We sample the log-volatilities, $\log(\sigma^2_{m1}), \ldots, \log(\sigma^2_{mT})$, and the parameters of the corresponding state equation (which involve the unconditional mean, the persistence parameter, and the variance of the shocks to the log-volatilities) using the efficient sampler devised in Kastner and Frühwirth-Schnatter (2014) and implemented in the R package stochvol (Kastner, 2016).

The following quantities are not estimated in an equation-by-equation manner:

11. **Sampling the latent factors $\{\boldsymbol{q}_t\}_{t=1}^{T}$.** The latent factors in $\boldsymbol{q}_t$ are simulated on a t-by-t basis. This can be achieved by estimating T separate regressions by regressing $\hat{\boldsymbol{y}}_t = \boldsymbol{\Sigma}_t^{-1/2}\left(\boldsymbol{y}_t - \sum_{p=1}^{P}(\boldsymbol{A}_p + \boldsymbol{B}_{pt})\boldsymbol{y}_{t-p}\right)$ on $\hat{\boldsymbol{X}}_t = \boldsymbol{\Sigma}_t^{-1/2}\boldsymbol{\Gamma}$ for all t. The corresponding time t posterior of $\boldsymbol{q}_t$ is Gaussian:

$$\boldsymbol{q}_t|\bullet \sim \mathcal{N}(\overline{\boldsymbol{q}}_t, \overline{\boldsymbol{V}}_{q_t}),$$

19

with covariance matrix and mean vector given by, respectively:

$$\overline{\boldsymbol{V}}_{q_t} = \left(\hat{\boldsymbol{X}}'_t \hat{\boldsymbol{X}}_t + \boldsymbol{R}_t^{-1} \right)^{-1} \quad \text{and} \quad \overline{\boldsymbol{q}}_t = \overline{\boldsymbol{V}}_{q_t} \hat{\boldsymbol{X}}'_t \hat{\boldsymbol{y}}_t.$$

12. **Sampling the hyperparameters of the Horseshoe associated with $\boldsymbol{\Gamma}$.** The Horseshoe shrinkage parameters on the factor loadings in $\boldsymbol{\Gamma}$ are simulated analogously to Step 8 of the algorithm. The main difference, however, is related to the fact that we specify Q_q global shrinkage parameters (one for each column) and the corresponding full conditional posterior for $\varpi_{\Gamma_j}^2$ needs to be adjusted by summing only over the relevant local scales and parameters associated with and in the jth column of $\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}_j$.

Notice that steps (1) to (3) yield a draw from $p(\{\mathcal{T}_{mq}^\beta, \boldsymbol{\mu}_{mq}, \boldsymbol{\lambda}_{mq}\}_{q=1}^{Q_\beta} | \bullet_{/\boldsymbol{\beta}_{mt}})$ where the notation $\bullet_{/\boldsymbol{\beta}_{mt}}$ indicates the remaining model parameters except the TVPs and the data. The TVPs are then simulated from $p(\{\boldsymbol{\beta}_{mt}\}_{t=1}^T | \bullet)$ where $\bullet$ means all other model parameters, latent quantities and the data. This step differs from the one used in Deshpande et al. (2020) since we improve mixing by integrating out the TVPs. In principle, the loadings and trees can also be sampled conditionally on the TVPs but in cases where the loadings are very small substantial mixing issues arise.

We repeat this algorithm $15,000$ times and discard the first $5,000$ draws as burn-in.[8] From a computational perspective, this algorithm is quite efficient. This is because the sampling step associated with the TVPs can be sped up enormously by exploiting the fact that $\boldsymbol{W}'_m \boldsymbol{W}_m$ is a block-diagonal matrix of rank T.

4 Empirical application: Modeling inflation

4.1 Data overview and specification choices

We use the quarterly version of the McCracken and Ng (2016) data set and focus on a sample ranging from 1975:Q1 to 2019:Q4. In our empirical work, we aim to investigate

---

[8]To obtain $15,000$ draws, the actual computation time is about 124 minutes, based on a MacBook Pro with an M1 8-core processor.

how business cycle shocks impact a range of different price measures and whether these dynamic reactions depend on the effect modifiers. To this end, we follow Del Negro et al. (2020) and estimate medium-sized VAR models that are rich in wage, price, and labor market measures. We consider $M = 12$ endogenous variables, where $\boldsymbol{y}_t$ includes output growth, employment, unemployment, average weekly hours worked, personal consumption expenditure (PCE) inflation, PCE inflation excluding food and energy, (core) consumer price inflation, the GDP deflator, wage inflation, the federal funds rate, and ten-year government bond yields to capture movements in treasury markets. But unlike Del Negro et al. (2020), we allow for nonlinear relationships between these variables and for these effects to vary over time. Del Negro et al. (2020) accommodate temporal change only, by simply estimating their linear VAR model over two non-overlapping samples.

As effect modifiers in $\boldsymbol{z}_t$, we consider five indicators that may affect the TVPs, and in turn the impulse response functions, in a nonlinear manner. Specifically, we consider the old-age dependency ratio, a financial globalization indicator, the (lagged) ex-post real rate, a binary recession indicator (taken from the NBER), and the economic uncertainty index proposed in Jurado et al. (2015). Secular stagnation factors, such as a boost in financial globalization, the rising old-age dependency ratio, and a declining real rate, may affect the dynamics of business cycle phases in a nonlinear manner (Jones, Forthcoming). These factors have also been identified as one cause of the flattening of the Phillips curve (Forbes, 2019; Forbes et al., 2021). The last two effect modifiers allow for possible structural breaks in recessionary and high uncertainty periods (see, for example, Aastveit et al., 2017; Alessandri and Mumtaz, 2019).

Some of the effect modifiers are clearly endogenous and should depend on the other quantities of our model. This does not cause any issues for the validity of our econometric approach. However, when we focus on impulse responses it has the implication that $\boldsymbol{z}_t$ is not allowed to react to changes in $\boldsymbol{y}_t$. As discussed in Section 2, this is an assumption made for the sake of interpretability. The main implication is that impulse responses can be understood as being conditional on $\boldsymbol{z}_t$ remaining at the current level over the impulse response horizon. Since we are going to construct "scenarios," based

on assumptions about how $\boldsymbol{z}_t$ behaves, this restriction can be interpreted as similar in nature to conditional forecasts when the restricted variables are not located in $\boldsymbol{y}_t$ but in $\boldsymbol{z}_t$. If the researcher wishes to relax these assumptions, they can set up auxiliary models for $\boldsymbol{z}_t$, such that $\boldsymbol{z}_t$ is again a function of $\boldsymbol{y}_t$.

Table B.1 in the Online Appendix provides additional information on the time series and associated data transformations used. All models we consider in this paper feature $p = 5$ lags. In Sub-section C.2 we assess how different model features impact model fit and compare our proposed model to standard models in the literature. This analysis evidences that our model generally captures the data well, often improving upon competitors commonly used in the literature. Based on the results in Table C.1, we use the model that sets $Q_\beta = 25$, $S_\beta = 1$, $Q_q = 3$, and $S_q = 250$.

4.2 Some features of our estimated model

In this sub-section we consider what is driving the time variation in the VAR coefficients in our TVP-BART model with FHB. Fig. 1(a) shows a heatmap of the total share of time-variation of $\boldsymbol{\beta}_{mt}$ explained by the nonlinear factors $F_m(\boldsymbol{z}_t)$ across equations $m = 1, \dots, M$. This quantity, closely related to the familiar R^2, is computed as follows:

$$\text{diag}\left(\frac{\sum_{q=1}^{Q_\beta} \boldsymbol{\lambda}_{mq}\text{Var}(g(\boldsymbol{z}_t|\mathcal{T}_{mq}^\beta, \boldsymbol{\mu}_{mq})\boldsymbol{\lambda}_{mq}'}{\sum_{q=1}^{Q_\beta} \boldsymbol{\lambda}_{mq}\text{Var}(g(\boldsymbol{z}_t|\mathcal{T}_{mq}^\beta, \boldsymbol{\mu}_{mq})\boldsymbol{\lambda}_{mq}' + \boldsymbol{V}_m}\right),$$

with $\text{Var}(g(\boldsymbol{z}_t|\mathcal{T}_{mq}^\beta, \boldsymbol{\mu}_{mq})$ denoting the empirical variance of the function g. Dark red values indicate that a given TVP is driven almost exclusively by $F_m(\boldsymbol{z}_t)$, whereas white values suggest that most of the variation is driven by idiosyncratic movements in the TVPs.

Panel (b) of Fig. 1 displays a heatmap of posterior means of the number of tree splits induced by one of the effect modifiers in $\boldsymbol{z}_t$ across coefficients and equations. This serves as a way to assess the relative importance of different effect modifiers in shaping the coefficient dynamics over time.

Starting with panel (a) of the figure, we see that the explanatory power of the TVP

factors varies substantially across equations (and also across variables). While we find that TVPs in the interest rate and CPI core equations are strongly shaped by the effect modifiers, this share is considerably lower for the other equations. With two exceptions (PCETCPI and GDPCTPI), the shares are, however, sizable and often above 50 percent. Turning to PCETCPI and GDPCTPI, the effect modifiers explain a rather small amount of variation. Interestingly, for labor market quantities (EMPL, UNRATE, AWH) and real GDP we also find that the intercept (which determines the unconditional mean of the model) is strongly influenced by different effect modifiers. This indicates that long-run properties of these time series depend on covariates that may be interpreted as capturing structural change in the macroeconomy.

Fig. 1: Explaining the dynamics in coefficients of the conditional mean.
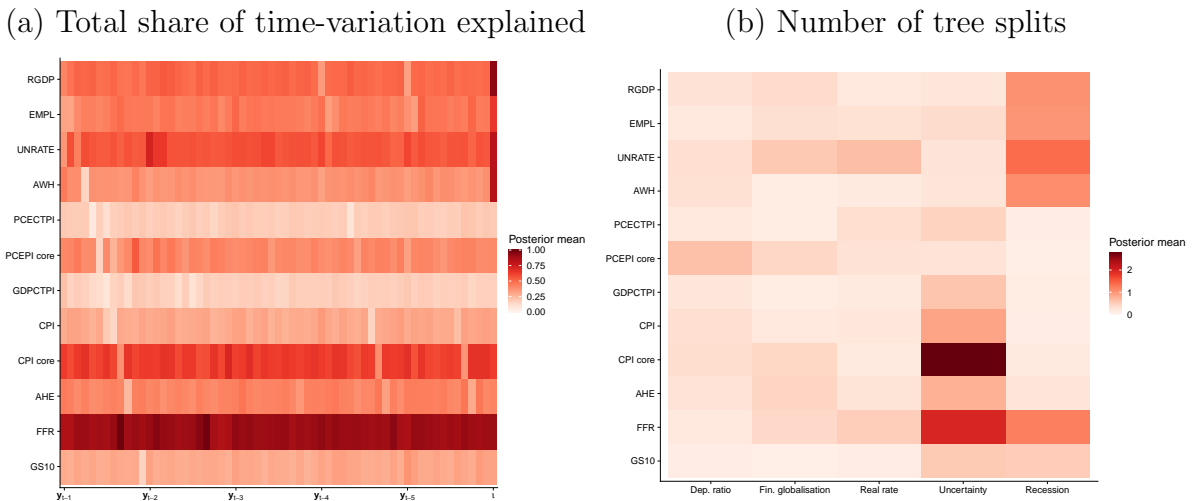
(a) Total share of time-variation explained   (b) Number of tree splits

Notes: Main specification with $Q_\beta = 25$ ($S_\beta = 1$) for the conditional mean and $Q_q = 3$ ($S_q = 250$) for the conditional variance-covariances. Panel (a) shows the posterior mean of the total share of time-variation explained for each coefficient. Vertical axis: endogenous variables. Front axis: coefficients related to the lags of $\boldsymbol{y}_t$. Panel (b) shows the posterior mean of the number of tree splits triggered by a certain effect modifier. For each effect modifier the number of splits is summed over the $Q_\beta (= 25)$ factors for the conditional mean. Vertical axis: endogenous variables. Front axis: effect modifiers.

Focusing on panel (b) of the figure provides additional insights. First, the old-age dependency ratio, financial globalization, and the real rate play only a limited role in explaining parameter dynamics. Second, for several variables we find that uncertainty shapes TVP dynamics. Among these are coefficients in the CPI and CPI core equations, the short-term interest rate equation, and the ten-year government bond yield. Third, for other variables such as output, employment, and the unemployment rate, we observe that uncertainty plays a more limited role. However, in these equations we instead find

that the NBER's recession indicator is frequently included in the splitting rules.

4.3 Capturing business cycle shocks

One of the main advantages of our nonparametric model is that, conditional on knowing the TVPs and error covariances, the model is a standard linear TVP-VAR model. Hence structural analysis, using identified impulse responses, can be readily carried out. In principle, an economist's preferred identification strategy based on, for example, sign restrictions (Benati, 2008), zero impact restrictions (Primiceri, 2005; Koop et al., 2009), or long-run restrictions can be implemented within our TVP-BART framework.

In this application, we focus on the question of how adverse business cycle shocks impact a set of inflation measures. To do so, we exploit the factor structure on the reduced-form VAR shocks to identify a business cycle shock (for related identification approaches, see Korobilis, 2022; Chan et al., 2022). As emphasized by Gorodnichenko (2005), in VAR models like ours where the number of variables is relatively large (we have $M = 12$) it can facilitate structural interpretation to have fewer structural shocks than M. In the next step, we trace out the dynamic evolution of our inflation measures to such a business cycle shock.

One can decompose, as in Eq. (2), the reduced-form VAR shocks into a factor component and an idiosyncratic measurement-error component (both of which are independent) under standard conditions (see, for example, Anderson and Rubin, 1956; Frühwirth-Schnatter and Lopes, 2018; Kaufmann and Schumacher, 2019).[9] Absent heteroskedasticity, the resulting factors still have no economic interpretation and thus additional structure is required to identify the shocks, given that the factors and the factor loadings can be rotated by any random orthogonal matrix. But, given the heteroskedasticity in $\boldsymbol{R}_t$, we can follow Chan et al. (2022) and identify, up to sign and scale, a business cycle shock as that factor (shock) that explains the largest amount of variation in innovations to output and unemployment variations during recessionary periods (as identified by the NBER). This identification strategy resembles the one proposed in Bianchi et al. (2023). They

---

[9]These conditions relate to the number of factors being smaller then the Ledermann bound and the number of non-zero elements in $\boldsymbol{\Gamma}$ being sufficiently large so that the decomposition in Eq. (2) is unique.

identify business cycle shocks by searching for linear combinations of the reduced-form shocks of a trend-cycle VAR so as to maximize the amount of variation in unemployment or cyclical output.[10]

Specifically, our business cycle shock is obtained by computing:

$$\boldsymbol{\zeta}_{jt} = \text{diag}\left(\frac{r_{jt}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j'}{\boldsymbol{\Gamma}\boldsymbol{R}_t\boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_t}\right), \tag{9}$$

for all j and finding that factor that maximizes the variances explained for real GDP and the unemployment rate during recessionary episodes. This yields, for each MCMC draw, a factor that can be interpreted as a business cycle shock. To point-identify the sign of the factors and the associated loadings, we normalize the factors and loadings to identify the business cycle shock as having a negative impact effect on output growth and a positive impact effect on unemployment.

In summary, we identify the business cycle shock and the associated impulse responses via the following steps:

1. We identify the business cycle factor by finding the factor that explains the largest amount of variation in output growth and unemployment during NBER-defined recessions; see Eq. (9). Without loss of generality, let us assume that this factor is the jth element of $\boldsymbol{q}_t$, q_{jt}, and the corresponding loadings are $\boldsymbol{\gamma}_j$.

2. We compute the responses to a unit increase in q_{jt}. The impact reaction of y_t is given by the loadings $\boldsymbol{\gamma}_j$ and higher order impulse responses are computed using standard recursions based on the companion form of the VAR at time t.

3. To identify whether the shock is contractionary or expansionary (that is, the sign of $\boldsymbol{\gamma}_j$), we check the impact reaction for real output growth and the unemployment rate. If the former decreases and the latter increases, we label the shock as being contractionary.

4. Finally, to anchor the magnitude of the shock, we normalize the impact responses such

---

[10]Alternative approaches to identify business cycle shocks are proposed in Del Negro et al. (2020) and Angeletos et al. (2020).

that real output growth declines by one standard deviation on average.

These steps yield partial identification, implying that the business cycle shock is uniquely identified whereas the remaining $(Q_q - 1)$ factors (and the associated columns in $\mathbf{\Gamma}$) are left unrestricted. This identification approach is related to ones developed in recent papers (Korobilis, 2022; Chan et al., 2022) which advocate using sign restrictions on the factor loadings to pin down a shock of interest. But our approach differs in the sense that we solve the column switching problem (which is required to attach an economic meaning to the different factors) through a narrative approach that builds on the notion that business cycle shocks are the ones that determine the largest amount of variation in real activity quantities during recessions. Our approach could easily be combined with sign-restricted factor stochastic volatility models, by introducing certain restrictions on the prior associated with $\mathbf{\Gamma}$.

Fig. 2: Shock labeling. Share of the variance of business cycle variables explained by factors over time.
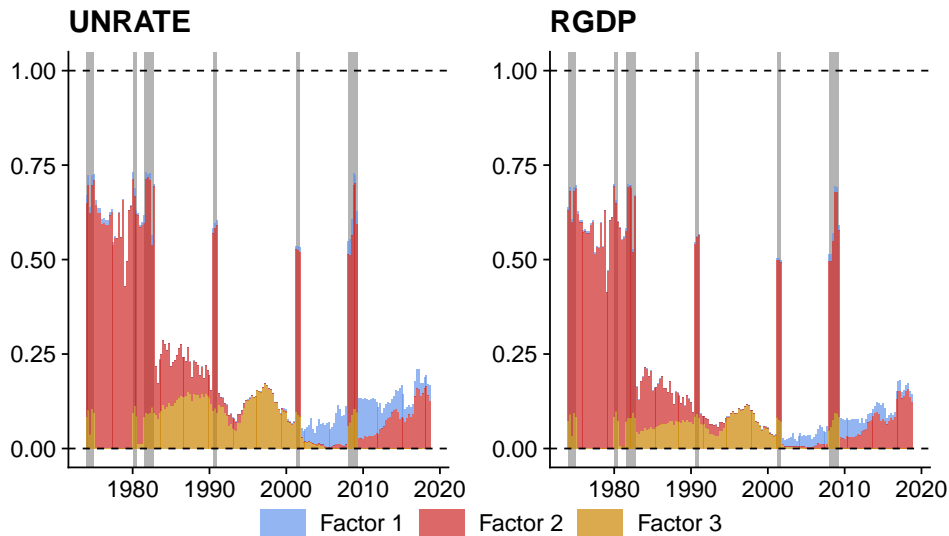
Notes: Main specification with $Q_\beta = 25$ ($S_\beta = 1$) for the conditional mean and $Q_q = 3$ ($S_q = 250$) for the conditional variance-covariances. This figure shows the posterior mean of the share of the variance of business cycle variables explained by factors over time (by focusing on the diagonal elements of $\mathbf{\Omega}_t$ related to unemployment and real output growth). We then refer to the factor that accounts for the maximal volatility of real output growth and unemployment as our business cycle shock. Vertical axis: share explained by each factor. Front axis: quarters.

Fig. 2 plots the posterior mean of the proportion of the variation, $\boldsymbol{\zeta}_{jt}$, in the unemployment rate and in output growth explained by the three factors over time. This figure shows that the second factor explains the largest amount of variation in the early part of the sample (until the twin recession of the early 1980s) and during all recessions in our

sample. During recessions, this factor explains close to 70 percent of the variation in the reduced-form shocks to both unemployment and output growth.

4.4 Impulse responses to a business cycle shock

In this sub-section, we look at the dynamic effects of our business cycle shock. Impulse response functions are computed by shocking the business cycle (second) factor and tracing out the dynamic reactions of y_{t+h} for $h = 1, \ldots, 16$.

Since our model features TVPs, the impulse responses can be computed at each point in time. This gives us a posterior distribution over T period-specific IRFs, a statistical object that is difficult to visualize. To aid exposition, we start our analysis by considering average impulse responses. These are obtained by averaging the time-specific impulse responses over time and are depicted in Fig. 3.
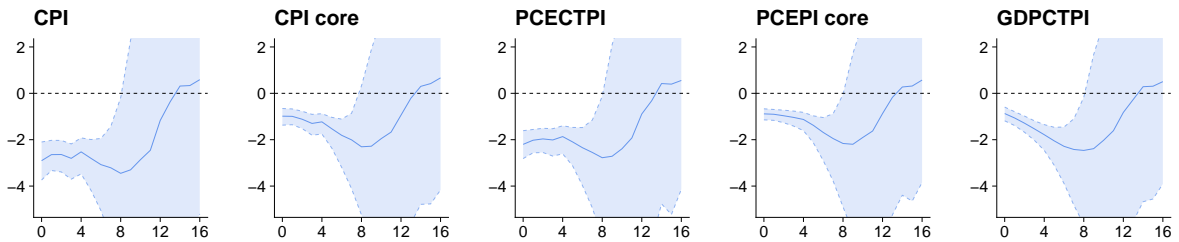
Fig. 3: Average impulse response over horizons.
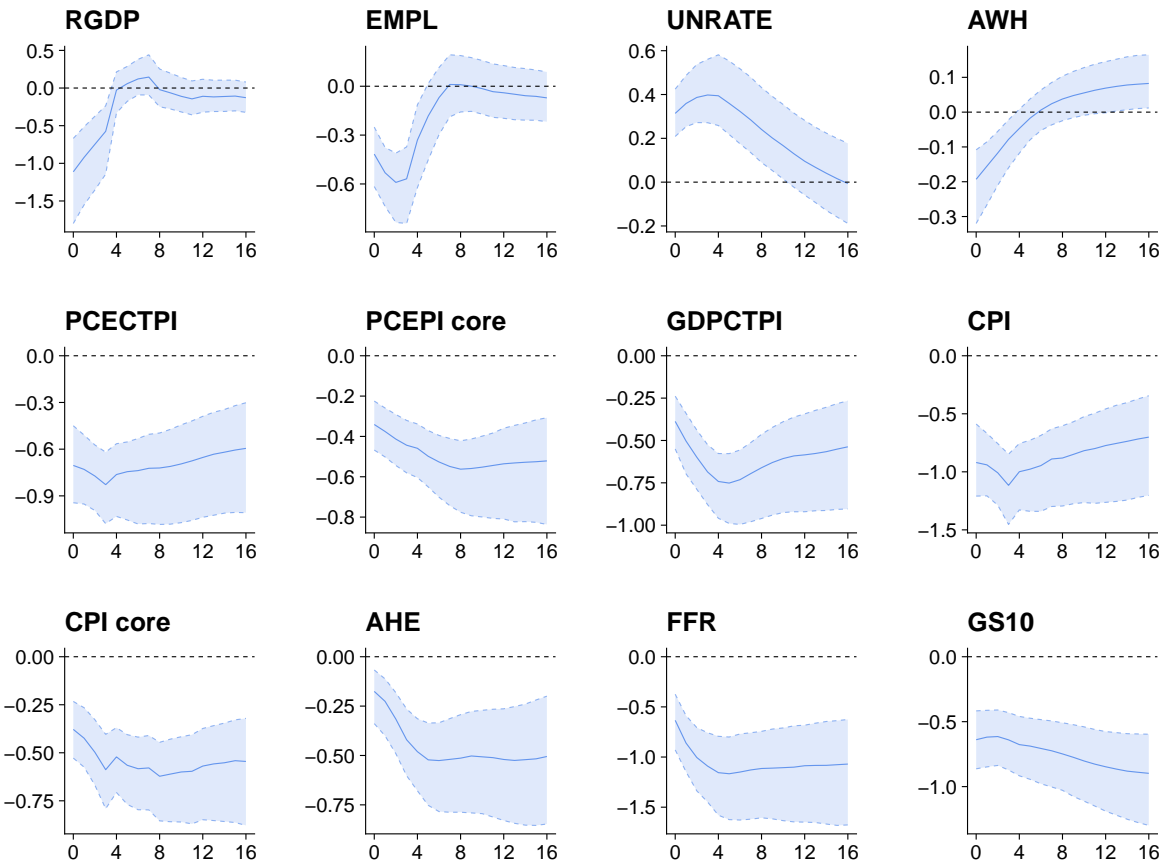
Notes: Impulse responses to a negative business cycle shock, averaged over time. Blue solid lines denote the posterior median, blue dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the blue shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Panels: endogenous variables. Vertical axis: impulse responses. Front axis: horizons (in quarters).

Fig. 3 shows that, averaged over time, a contractionary business cycle shock leads to unemployment rising and inflation (including core and wage inflation) falling. The dynamic effects on the different inflation measures are similar, but long-lasting. Like the main business shock of Angeletos et al. (2020), the peak effect of our business cycle shock on the real variables also occurs within a year or two. Specifically, we observe that output and employment decline while the unemployment rate increases. Real GDP growth reacts rapidly by declining by around one percentage point on impact. For employment growth, the peak effect materializes after about three quarters. The unemployment rate quickly increases and displays a peak reaction of around 0.5 percentage points after around one year. These reactions are largely consistent (both in terms of shape and size) with the ones reported in Bianchi et al. (2023).

Fig. 4: Average Phillips curve multipliers over horizons.

Notes: Philips curve multipliers based on a negative business cycle shock, averaged over time. Blue solid lines denote the posterior median, blue dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the blue shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Panels: price indices. Vertical axis: impulse responses. Front axis: horizons (in quarters).

When we consider the reactions of our different inflation measures, we find that prices decline on impact. This reduction appears to be quite persistent. As we will show below (see Fig. 5), this persistent reaction of prices is mainly driven by strong and persistent declines of inflation up to the early 1990s. These results are consistent with the existence of a negatively sloped Phillips curve — at least on average through the 1975:Q1 to 2019:Q4 period.

To hone in on this relationship between inflation and unemployment, we normalize the IRFs of the different price measures by the IRFs of the unemployment rate. Barnichon and Mesters (2021) call this quantity the Phillips curve multiplier. The multipliers are shown in Fig. 4. We again see that, on average over time, the Phillips curve multipliers are negative and statistically significant. Fig. 4 also reveals that these negative effects

persist for two to three years. And they vary by inflation measure. As we should expect, the Phillips curve is stronger for headline than core measures of inflation. The strongest effects on inflation are typically seen two years after the business cycle shock.

To understand to what degree averaging over time is masking temporal variations in the Phillips curve relationship, Fig. 5 plots, at the one-year-ahead horizon ($h = 4$), the impulse responses due to the contractionary business cycle shock at each point in time. The ability to identify and capture structural change of different forms is a key feature of our model. Fig. 5 reveals that there are indeed important temporal variations. The responses of, in particular, the headline inflation measures become more muted over time. Focusing in on the effects on CPI inflation, we see that the business cycle shock lowers inflation significantly through the 1970s and 1980s. But the responses thereafter are more muted. They become increasingly muted as we look to the period after the global financial crisis. Interestingly, evidencing a clear nonlinearity, there is a strong negative effect on inflation during the recessionary period associated with the global financial crisis itself. Our findings therefore provide ex-post justification for the decision by Del Negro et al. (2020) to estimate their VAR model, designed to understand the Phillips curve, on samples before and after 1990. But our results also reveal important temporal instabilities and changes within these two periods that are lost by simple sample-slit or indeed rolling regressions as also often used in the literature.

Turning to the effects on unemployment, again consistent with Del Negro et al. (2020), Fig. 5 shows that the response of unemployment to a business cycle shock becomes more persistent over time. This is consistent with economic expansions lasting longer in more recent decades. But Fig. 5 adds texture to this narrative by revealing that recessionary periods, except for 2008-9, are marked by especially strong responses.

Bringing together the price inflation and unemployment responses, we conclude that the sensitivity of price inflation to unemployment has weakened markedly since 1990. The response of wage inflation to the the business cycle shock is weaker throughout the sample. This casts doubt on the view (see, for example, Knotek and Zaman, 2014; Hooper et al., 2020) that the Phillips curve is stronger for wage than for price inflation. Since

the 1990s the impulse responses for wage inflation and price (CPI) inflation look broadly similar; see Fig. 5. This includes evidence that wage as well as price inflation did decline in response to a business cycle shock during the Great Recession, with prices declining by more than wages.

4.5   Scenario analysis to assess the channels of time variation

One key feature of our model is that it allows us to link the time variation in the parameters (and thus functions thereof such as IRFs) to the effect modifiers in z_t. Since z_t influences the TVPs using a nonparametric model, it is difficult to clearly answer how changes in z_t impact the TVPs. Since our interest centers on the implied IRFs, we can, however, carry out simulations that show how the dynamic responses to a business cycle shock change as we vary z_t. The results of this exercise are shown in Fig. 6. This figure depicts the price responses in the rows of the panel and in the columns shows different assumptions on z_t.

To analyze whether IRFs differ in expansion and recessions, we set the NBER recession indicator to zero (that is, we assume that the economy is in an expansion) or to one (that is, we assume that the economy is in a recession). Based on this, we vary one of the effect modifiers while setting the remaining effect modifiers to some pre-specified value. This pre-specified value is either the average value over the period 1975 to 1985 (which are the blue-shaded IRFs in the figure) or the period after 2010 (which are the red-shaded IRFs). This allows us to capture the general macroeconomic environment in the respective time periods. This gives us four overall combinations for the IRFs. We consider how prices react in expansions and recessions and whether there are discernible differences in the transmission of business cycle shocks in these two regimes. Based on one of these four general scenarios, we set each effect modifier (for example, the dependency ratio, financial globalization, the real rate, and uncertainty) equal to different sample quantiles and then compute the implied price IRFs. This provides a detailed picture on how impulse responses depend on the effect modifiers.

Fig. 6 confirms that the headline (non-core) inflation measures were more strongly

Fig. 5: Impulse responses over time at the one-year-ahead horizon ($h = 4$).
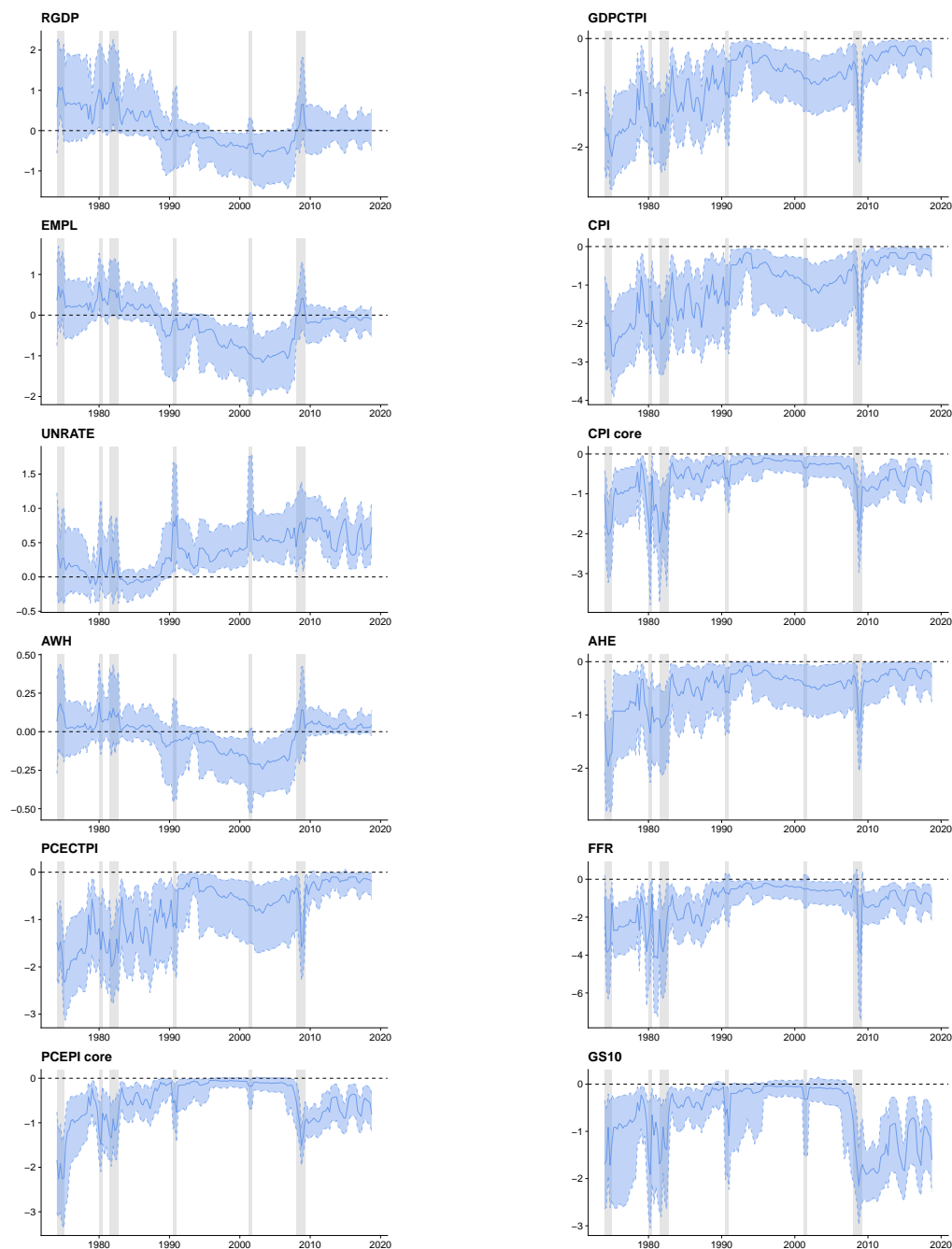
Notes: Time-specific impulse responses to a negative business cycle shock. Blue solid lines denote the posterior median, blue dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the blue shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Panels: endogenous variables. Vertical axis: impulse responses. Front axis: periods (in quarters).

Fig. 6: Impulse response of prices for different scenarios at the one-year-ahead horizon ($h = 4$).
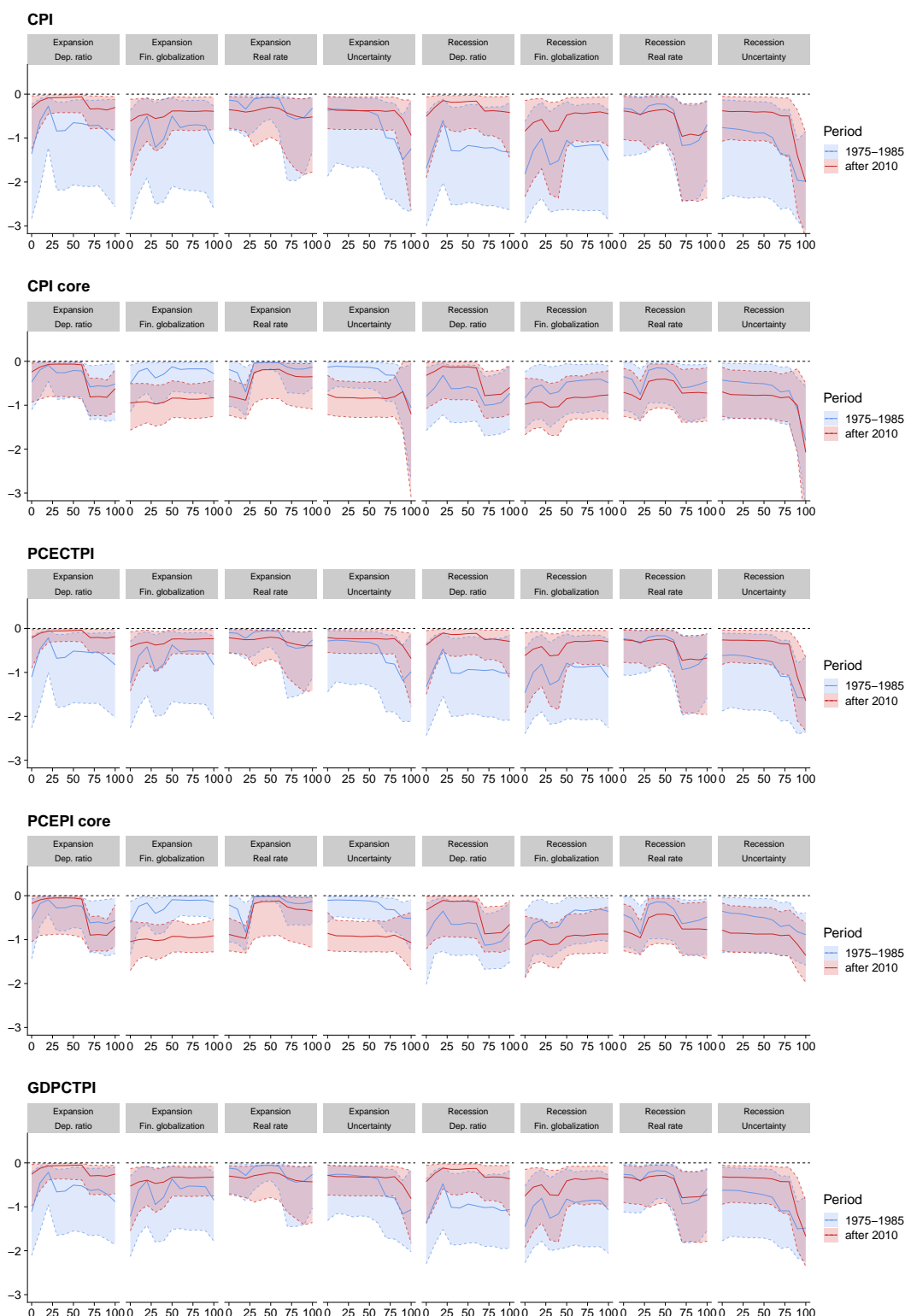
Notes: Impulse responses to a negative business cycle shock by partially varying the effect modifiers. For example, the top-left panel refers to the responses across percentiles of the dependency ratio while assuming an expansion state and setting the remaining effect modifiers (that is, financial globalization, real interest rate, and uncertainty) either to the mean of the subsample of periods 1975:Q1 to 1984:Q4 (colored in blue) or to the mean of the subsample of periods from 2010:Q1 to 2019:Q4 (colored in red). Colored solid lines denote the posterior median, colored dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the colored shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Vertical panels: price indices. Horizontal panels: effect modifiers for an expansion and a recession state. Vertical axis: impulse responses. Front axis: percentiles in % where 0% (100%) denotes the minimum (maximum) value.

affected by business cycle shocks before 1985. The most striking nonlinearity for the effect modifiers is seen with respect to uncertainty. As uncertainty increases beyond its 75th percentile, we see much stronger negative effects on all the inflation measures, including with post-2010 data. This effect is especially pronounced during recessionary periods. This all supports a view that the Phillips curve remains alive and well during times of recession and greater-than-average uncertainty, events that empirically tend to co-exist. This is consistent with theories of the financial accelerator, suggesting that shocks have amplified effects in recessions.

5   Conclusion

In this paper, we have developed a nonparametric model that uses Bayesian additive regression trees (BART) methods to allow for change of an unknown form in both the conditional means and variances of a multivariate time series model (a VAR). Unlike existing nonparametric approaches, interpretation and macroeconomic inference including structural analysis is easier, since, as the model gives a nonparametric treatment to the parameters rather than the variables, it remains conditionally linear in the mean. An additional novel feature is that the new model allows for nonparametric factor structures for parameters in the conditional means and variances, thus reducing the number of nonparametric functions to estimate and ensuring parsimony.

In an empirical exercise we show how the proposed nonparametric VAR model contributes to our understanding of the time-varying nature of the Phillips curve. Inflation has become considerably less sensitive to business cycle shocks, in particular since 1990. However, the flexible nonlinear features of the model show that the effects on inflation remain strong when uncertainty rises to high levels.

References

AASTVEIT, K. A., G. J. NATVIK, AND S. SOLA (2017): "Economic uncertainty and the influence of monetary policy," *Journal of International Money and Finance*, 76, 50–67.

ALESSANDRI, P., AND H. MUMTAZ (2019): "Financial regimes and uncertainty shocks," *Journal of Monetary Economics*, 101, 31–46.

ANDERSON, T. W., AND H. RUBIN (1956): "Statistical inference in factor models," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability: Held at the Statistical Labo-*

ratory, University of California, December, 1954, July and August, 1955, vol. 1, p. 111. University of California Press.

ANGELETOS, G.-M., F. COLLARD, AND H. DELLAS (2020): "Business-Cycle Anatomy," *American Economic Review*, 110(10), 3030–70.

BAI, J., AND S. NG (2007): "Determining the number of primitive shocks in factor models," *Journal of Business & Economic Statistics*, 25(1), 52–60.

BARNICHON, R., AND G. MESTERS (2021): "The Phillips multiplier," *Journal of Monetary Economics*, 117, 689–705.

BELMONTE, M. A., G. KOOP, AND D. KOROBILIS (2014): "Hierarchical shrinkage in time-varying parameter models," *Journal of Forecasting*, 33(1), 80–94.

BENATI, L. (2008): "The "Great Moderation" in the United Kingdom," *Journal of Money, Credit and Banking*, 40(1), 121–147.

BIANCHI, F., G. NICOLÒ, AND D. SONG (2023): "Inflation and Real Activity over the Business Cycle," Working Paper 31075, National Bureau of Economic Research.

BITTO, A., AND S. FRÜHWIRTH-SCHNATTER (2019): "Achieving shrinkage in a time-varying parameter model framework," *Journal of Econometrics*, 210(1), 75–97.

CAGGIANO, G., E. CASTELNUOVO, AND G. PELLEGRINO (2017): "Estimating the real effects of uncertainty shocks at the zero lower bound," *European Economic Review*, 100, 257–272.

CHAN, J. C. (Forthcoming): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*.

CHAN, J. C., E. EISENSTAT, AND R. W. STRACHAN (2020): "Reducing the state space dimension in a large TVP-VAR," *Journal of Econometrics*, 218(1), 105–118.

CHAN, J. C., E. EISENSTAT, AND X. YU (2022): "Large Bayesian VARs with factor stochastic volatility: Identification, order invariance and structural analysis," Discussion paper, arXiv preprint arXiv:2207.03988.

CHIPMAN, H. A., E. I. GEORGE, AND R. E. McCULLOCH (1998): "Bayesian CART model search," *Journal of the American Statistical Association*, 93(443), 935–948.

——— (2010): "BART: Bayesian additive regression trees," *The Annals of Applied Statistics*, 4(1), 266–298.

CLARK, T. E., F. HUBER, G. KOOP, M. MARCELLINO, AND M. PFARRHOFER (Forthcoming): "Tail forecasting with multivariate Bayesian additive regression trees," *International Economic Review*.

COGLEY, T., AND T. J. SARGENT (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8(2), 262 – 302.

COULOMBE, P. G. (2020): "The macroeconomy as a random forest," Discussion paper, arXiv preprint arXiv:2006.12724.

D'AGOSTINO, A., L. GAMBETTI, AND D. GIANNONE (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28(1), 82–101.

DANGL, T., AND M. HALLING (2012): "Predictive regressions with time-varying coefficients," *Journal of Financial Economics*, 106(1), 157–181.

DEL NEGRO, M., M. LENZA, G. E. PRIMICERI, AND A. TAMBALOTTI (2020): "What's up with the Phillips Curve?," Working Paper 27003, National Bureau of Economic Research.

DESHPANDE, S. K., R. BAI, C. BALOCCHI, J. E. STARLING, AND J. WEISS (2020): "VCBART: Bayesian trees for varying coefficients," Discussion paper, arXiv preprint arXiv:2003.06416.

FISCHER, M. M., N. HAUZENBERGER, F. HUBER, AND M. PFARRHOFER (2023): "General Bayesian time-varying parameter vector autoregressions for modeling government bond yields," *Journal of Applied Econometrics*, 38(1), 69–87.

FORBES, K. (2019): "Inflation dynamics: Dead, dormant, or determined abroad?," Working paper 26496, National Bureau of Economic Research.

FORBES, K., J. GAGNON, AND C. G. COLLINS (2021): "Low inflation bends the Phillips curve around the world," Working paper 29323, National Bureau of Economic Research.

FRÜHWIRTH-SCHNATTER, S., AND H. F. LOPES (2018): "Sparse Bayesian factor analysis when the number of factors is unknown," Discussion paper, arXiv preprint arXiv:1804.04231.

FRÜHWIRTH-SCHNATTER, S., R. TÜCHLER, AND T. OTTER (2004): "Bayesian analysis of the heterogeneity model," *Journal of Business & Economic Statistics*, 22(1), 2–15.

FRÜHWIRTH-SCHNATTER, S., AND H. WAGNER (2010): "Stochastic model specification search for Gaussian and partial non-Gaussian state space models," *Journal of Econometrics*, 154(1), 85–100.

GELMAN, A., J. HWANG, AND A. VEHTARI (2014): "Understanding predictive information criteria for Bayesian models," *Statistics and Computing*, 24(6), 997–1016.

GORODNICHENKO, Y. (2005): "Reduced-Rank Identification of Structural Shocks in VARs," Macroeconomics 0512011, University Library of Munich, Germany.

HAUZENBERGER, N., F. HUBER, G. KOOP, AND L. ONORANTE (2022): "Fast and flexible Bayesian inference in time-varying parameter regression models," *Journal of Business & Economic Statistics*, 40(4), 1904–1918.

HAUZENBERGER, N., M. PFARRHOFER, AND A. STELZER (2021): "On the effectiveness of the European

Central Bank's conventional and unconventional policies under uncertainty," *Journal of Economic Behavior & Organization*, 191, 822–845.

HOOPER, P., F. S. MISHKIN, AND A. SUFI (2020): "Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating?," *Research in Economics*, 74(1), 26–62.

HUBER, F., G. KOOP, AND L. ONORANTE (2021): "Inducing sparsity and shrinkage in time-varying parameter models," *Journal of Business & Economic Statistics*, 39(3), 669–683.

HUBER, F., G. KOOP, L. ONORANTE, M. PFARRHOFER, AND J. SCHREINER (2023): "Nowcasting in a pandemic using non-parametric mixed frequency VARs," *Journal of Econometrics*, 232(1), 52–69.

HUBER, F., AND L. ROSSINI (2022): "Inference in Bayesian additive vector autoregressive tree models," *The Annals of Applied Statistics*, 16(1), 104–123.

HUBRICH, K., AND R. J. TETLOW (2015): "Financial stress and economic dynamics: The transmission of crises," *Journal of Monetary Economics*, 70, 100–115.

JONES, C. (Forthcoming): "Aging, secular stagnation and the business cycle," *The Review of Economics and Statistics*.

JURADO, K., S. C. LUDVIGSON, AND S. NG (2015): "Measuring uncertainty," *American Economic Review*, 105(3), 1177–1216.

KASTNER, G. (2016): "Dealing with stochastic volatility in time series using the R package stochvol," *Journal of Statistical Software*, 69(5), 1–30.

KASTNER, G., AND S. FRÜHWIRTH-SCHNATTER (2014): "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models," *Computational Statistics & Data Analysis*, 76, 408–423.

KASTNER, G., AND F. HUBER (2020): "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, 39(7), 1142–1165.

KAUFMANN, S. (2015): "K-state switching models with time-varying transition distributions—Does loan growth signal stronger effects of variables on inflation?," *Journal of Econometrics*, 187(1), 82–94.

KAUFMANN, S., AND C. SCHUMACHER (2019): "Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification," *Journal of Econometrics*, 210(1), 116–134.

KNOTEK, E., AND S. ZAMAN (2014): "On the relationships between wages, prices, and economic activity," *Federal Reserve Bank of Cleveland, Economic Commentary*.

KOOP, G., AND D. KOROBILIS (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198.

KOOP, G., R. LEON-GONZALEZ, AND R. W. STRACHAN (2009): "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33(4), 997–1017.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74(1), 119–147.

KOOP, G., AND S. M. POTTER (2007): "Estimation and forecasting in models with multiple breaks," *The Review of Economic Studies*, 74(3), 763–789.

KOROBILIS, D. (2021): "High-dimensional macroeconomic forecasting using message passing algorithms," *Journal of Business & Economic Statistics*, 39(2), 493–504.

——— (2022): "A new algorithm for structural restrictions in Bayesian vector autoregressions," *European Economic Review*, 148, 104241.

LAUBACH, T., AND J. C. WILLIAMS (2003): "Measuring the natural rate of interest," *Review of Economics and Statistics*, 85(4), 1063–1070.

MAKALIC, E., AND D. F. SCHMIDT (2015): "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23(1), 179–182.

MCCRACKEN, M. W., AND S. NG (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

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.

PRATOLA, M. T., H. A. CHIPMAN, E. I. GEORGE, AND R. E. MCCULLOCH (2020): "Heteroscedastic BART via multiplicative regression trees," *Journal of Computational and Graphical Statistics*, 29(2), 405–417.

PRIMICERI, G. (2005): "Time varying structural autoregressions and monetary policy," *The Review of Economic Studies*, 72(3), 821–852.

SIMS, C. A., AND T. ZHA (2006): "Were there regime switches in US monetary policy?," *American Economic Review*, 96(1), 54–81.

STEVANOVIC, D. (2016): "Common time variation of parameters in reduced-form macroeconomic models," *Studies in Nonlinear Dynamics & Econometrics*, 20(2), 159–183.

VEHTARI, A., A. GELMAN, AND J. GABRY (2017): "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, 27(5), 1413–1432.

WATANABE, S. (2010): "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research*, 11, 3571–3594.

——— (2013): "A widely applicable Bayesian information criterion," *Journal of Machine Learning Research*, 14, 867–897.

Online Appendices

Bayesian Modeling of TVP-VARs
Using Regression Trees

NIKO HAUZENBERGER,[1,2] FLORIAN HUBER,[1]
GARY KOOP,[2] and JAMES MITCHELL[3]

[1] *University of Salzburg*
[2] *University of Strathclyde*
[3] *Federal Reserve Bank of Cleveland*

A   Additional details about the prior

In the main body of the paper we have focused on the prior for the tree part of the model. The remaining priors are relatively standard in the literature.

For the time-invariant VAR coefficients ($\boldsymbol{A}$), the factor loadings associated with the static error factors ($\boldsymbol{\Gamma}$), and the factor loadings in the state equation ($\boldsymbol{\Lambda}_m$) we use a Horseshoe prior. Let ϕ_{ij} denote a generic coefficient. A general Horseshoe prior is given by:

$$\phi_{ij}|\varrho_{ij}\varpi \sim \mathcal{N}(0, \varrho_{ij}^2\varpi^2), \quad \varrho_{ij} \sim \mathcal{C}^+(0,1), \quad \varpi \sim \mathcal{C}^+(0,1),$$

where ϱ_{ij} is a coefficient-specific shrinkage parameter that forces each coefficient to zero and ϖ denotes an equation-specific shrinkage parameter that forces all elements in the respect coefficient matrix ($\boldsymbol{A}$, $\boldsymbol{\Lambda}_m$, or $\boldsymbol{\Gamma}$) to zero. This prior differs across parameter types in one important way. For the VAR coefficients, the global shrinkage parameter applies to all elements in $\boldsymbol{A}$. For the factor loadings, we estimate separate global shrinkage parameters for each column of $\boldsymbol{\Lambda}_m$ and $\boldsymbol{\Gamma}$, respectively. This enables us to determine the number of factors by forcing columns to zero, thus excluding a specific factor from the model.

B Data Appendix

Table B.1: Data description.

Label	Mnemonic	Description	Transformation
		(a) Endogenous variables $\boldsymbol{y}_t$	
RGDP	GDPC1	Real gross domestic product	y-o-y growth rate
EMPL	CE16OV	Employment	y-o-y growth rate
UNRATE	UNRATE	Unemployment rate	none
AWH	CES0600000007	Avg. weekly hours (production and nonsuperv. employees)	none
PCECTPI	PCECTPI	Chain-type price index: personal consumption expenditures	y-o-y growth rate
PCEPI core	PCEPILFE	Personal consumption expenditures excl. food and energy	y-o-y growth rate
GDPCTPI	GDPCTPI	Chain-type price index: gross domestic product	y-o-y growth rate
CPI	CPIAUCSL	Consumer price index (all items)	y-o-y growth rate
CPI core	CPILFESL	Core consumer price index (all items less food and energy)	y-o-y growth rate
AHE	CES0600000008	Avg. hourly earnings (production and nonsuperv. employees)	y-o-y growth rate
FFR	FEDFUNDS	Effective federal funds rate	none
GS10	GS10	10-year government bond yields	none
		(b) Exogenous effect modifiers $\boldsymbol{z}_t$	
Dep. ratio	–	Age dependency ratio: older dependents to working-age population	none
Fin. globalization	–	KOF financial globalization index	none
Real rate	–	Laubach and Williams (2003) estimate of the natural rate of interest	none
Recession	–	National Bureau of Economic Research (NBER) recession indicator	none
Uncertainty	–	Jurado et al. (2015) estimate of macroeconomic uncertainty	none

Notes: For the endogenous variables we rely on the McCracken and Ng (2016) data set, while for the exogenous effect modifiers we draw on different data sources. The age dependency ratio and the NBER recession indicator are obtained from the FRED database of the Federal Reserve Bank of St. Louis (fred.stlouisfed.org), the financial globalization indicator is downloaded from KOF Swiss Economic Institute at the ETH Zurich (kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html), the real rate estimate of Laubach and Williams (2003) from the database of the Federal Reserve Bank of New York (www.newyorkfed.org/research/policy/rstar), and the macroeconomic uncertainty from the personal webpage of Sydney C. Ludvigson (sydneyludvigson.com/macro-and-financial-uncertainty-indexes). Since both the age dependency ratio and the financial globalization indicator are available only on an annual basis, they are interpolated to the quarterly frequency using a Kalman smoother.

C Empirical Appendix

C.1 How BART can approximate TVPs: An Illustration

In this sub-section, we provide a simple illustration of how BART works when used for approximating TVPs. We use a single equation example, with y_t being quarterly US CPI inflation and x_t being the unemployment rate.[1] Hence, $M = K = 1$. Moreover, we set $Q_\beta = Q_q = S_q = 1$ and focus this illustration on the role the number of trees in the conditional mean (S_β) plays in modeling the TVPs. Thus, there is a single TVP which is the Phillips curve coefficient (that is, the coefficient on the unemployment rate). We compare our estimate of this using various choices for S_β to the estimates produced by a standard TVP regression, where the Phillips curve coefficient evolves according to a random walk. To aid in comparability of the Philips curve coefficient across models, every model includes an intercept that evolves according to a random walk.

---

[1]Details about the data used in this empirical illustration are provided in Section 5. The model is estimated using the MCMC methods described in Section 3.2 of the main paper.

Under these assumptions our model reduces to:

$$y_t = c_t + \beta_t x_t + (q_t + \varepsilon_t), \quad (q_t + \varepsilon_t) \sim \mathcal{N}(0, R(z_t) + \sigma_t^2),$$

$$\beta_t = F(z_t) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, v^2),$$

where c_t has a random walk state equation. The conventional TVP regression replaces the second equation with a random walk but otherwise is identical.

We illustrate our BART-based techniques with $z_t = t$, so that the only explanatory variable in BART is a deterministic trend. In this case, the splitting rules would divide the time periods into distinct regimes that would feature their own terminal node parameters. The resulting model can be interpreted as a regime switching model with an unknown number of regimes and a diagonal transition probability matrix. This specification is closely related to the TVP regression model developed in Hauzenberger et al. (2022), which uses sparse finite mixtures to model the time-variation in the coefficients.

Figure C.1 shows the single estimated tree we obtain when we set $S_\beta = 1$. The oval boxes in the bottom row of the figure contain estimated values of the Phillips curve coefficient produced by the tree (along with the percentage of observations which share each value). Since we have set $z_t = t$ the tree divides up observations into different time periods. Hence, we basically have a regime structure where different regimes have different Phillips curve coefficients. For instance, eight percent of the observations have a Phillips curve coefficient which is very negative (-0.67). Proceeding from the top of the tree, it can be seen that these observations occur in the interval $[1984, 1988)$.

The regime structure can also be seen in Fig. C.2, which plots different estimates of β_t, including the case above with $S_\beta = 1$. In particular, it plots the (smoothly evolving) estimate produced by the TVP regression with random walk evolution of the coefficients against estimates produced by a single tree, five trees $(S_\beta = 5)$, and many trees $(S_\beta = 150)$. All four lines in the figures are similar to one another, but it can be seen that adding more trees allows for more regimes and the fitted line produced by BART becomes smoother and tracks the random walk evolution of the parametric model more and more closely.

Fig. C.1: Estimated branching structure for the Phillips curve coefficient using a single tree.
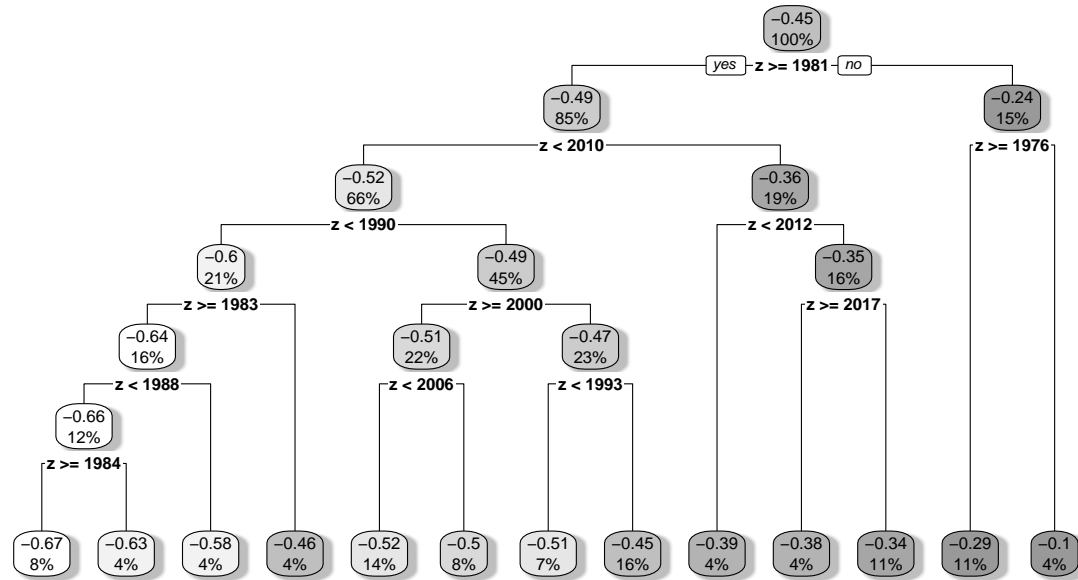
Notes: As the exogenous effect modifier (z), we use a simple linear time trend ranging from 1974:Q1 to 2019:Q4. Each oval box indicates the terminal node parameter of a particular branch and the share (in percent) of observations belonging to this branch. The splitting rules effectively decompose the time-varying coefficient into distinct regimes, featuring their own terminal node parameters.

Fig. C.2: Estimates of the time-varying Phillips curve coefficient using a single tree, five trees ($S_\beta = 5$), and many trees ($S_\beta = 150$).
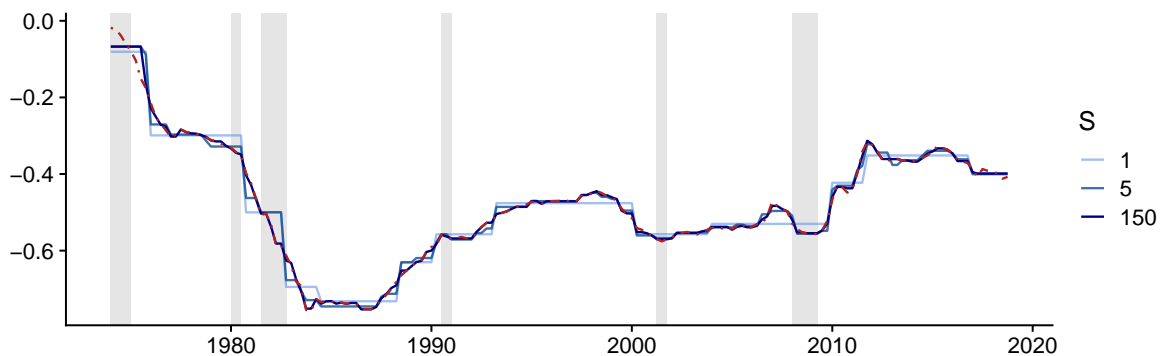
Notes: The blue lines indicate estimates of β_t for a varying number of trees, while the red dash-dotted line refers to estimates of β_t assuming random walk evolution of β_t. These models are estimated as special cases of our general model using the MCMC algorithm outlined in Sub-section 3.2.

C.2 Assessing model fit using information criteria

To assess model adequacy and decide on some parameters of the model, we focus on the widely applicable/Watanabe-Akaike information (WAIC, Watanabe, 2010, 2013). The WAIC is a generally applicable measure of model fit that, on the one hand, rewards model fit while, on the other hand, punishing model complexity. We opt for the WAIC due to its excellent characteristics when applied to tightly parameterized hierarchical models.[2] Following Gelman et al. (2014), the WAIC for model A is given by:

$$\text{WAIC}_A = -2 \left(\widehat{\text{lpd}}_A - \hat{p}_A \right).$$

Here, $\widehat{\text{lpd}}_A$ denotes an estimate of the log point-wise predictive density:

$$\widehat{\text{lpd}}_A = \sum_{t=1}^{T} \log \left(\frac{1}{S} \sum_{s=1}^{S} p_A(\boldsymbol{y}_t | \boldsymbol{\Theta}_A^{(s)}) \right),$$

where $\boldsymbol{\Theta}_A^{(s)}$ is generic notation that refers to the sth draw of the parameters and latent states from the full posterior distribution of model A. This term measures model fit. Model complexity is measured through $\hat{p}_A$. This quantity can be interpreted as the effective number of parameters and is given by the variance of the point-wise log likelihood across S draws from the full posterior distribution:

$$\hat{p}_A = \sum_{t=1}^{T} \text{Var} \left(\log p_A(\boldsymbol{y}_t | \{\boldsymbol{\Theta}_A^{(s)}\}_{s=1}^{S}) \right).$$

We compare our nonparametric TVP-FBART-FHB to alternatives that retain factor structures for the TVPs and multivariate stochastic volatility processes but are parametric.[3] Specifically, we compare our approach to the model proposed in Chan et al. (2020) using the factor stochastic model proposed in Kastner and Huber (2020). This model has a similar structure to ours, involving a factor structure both in the TVPs

---

[2]Gelman et al. (2014) and Vehtari et al. (2017) thoroughly study the characteristics of the WAIC.

[3]We do not compare our approach to TVP-VARs lacking such factor structures, such as the model of Primiceri (2005), since they would be heavily over-parameterized in models of our dimension. Thus, all of the models in our comparison adopt factor structures to ensure parsimony.

and the multivariate stochastic volatility process. We assume random walk behavior for the factor driving both the TVPs and the stochastic volatility process. We use the abbreviation TVP-FRW-FSV for this model. We emphasize that our approach differs from TVP-FRW-FSV only in modeling time variation nonparametrically, instead of with random walks. We also consider a constant-coefficient VAR with FSV.

In Table C.1 we compute WAICs for TVP-FBART-FHB and TVP-FRW-FSV for different numbers of factors (Q_β and Q_q) as well as, for the BART-based approaches, the number of trees driving the TVPs (S_β). This lets us pin down a preferred model specification and investigate the relationship between Q_β and S_β. The trade-off between Q_β and S_β is potentially interesting. In principle, setting Q_β to a large value and S_β to a small value leads to a model which is closely related to a standard BART specification for TVP-VARs (that is, many factors driving the TVPs but each is relatively simple, involving a small number of trees). In contrast, a model that sets Q_β to a small value and S_β to a large one implies the TVPs are driven by a small number of factors, but these factors are potentially very complicated involving a large number of trees. We investigate this relationship by considering different combinations of Q_β and S_β that reflect a range of cases. Since this relationship is less relevant in the lower dimensional error covariance matrix and we have found more robustness to choice of number of trees (provided it is not too small), we follow the literature on heteroBART and set $S_q = 250$.

Table C.1 suggests that performance of the TVP-FBART-FSV model strongly depends on the choices of Q_β, Q_q, and S_β. The best performance is obtained by setting both the number of factors in the conditional mean to be large (but with a small number of trees) and the number of factors in the FHB process to be small. Setting the number of factors in the FHB part of the model to be large leads to a deterioration in performance, regardless of the choices made relating to the FHB model. This suggests a fair degree of common volatility, but less commonality for the conditional mean coefficients.

We do see some trade-off between the number of factors and the number of trees in the FBART part of the model. Models with many factors and few trees perform best, but models with fewer factors and more trees perform nearly as well. The worst performance

Table C.1: Model comparison based on relative WAICs.

| VAR specification	Conditional mean			Conditional variances Number of factors Q_q			
	Q_β	S_β		1	3	6	12
TVP-BART-FHB	1	100		0.80	0.99	1.52	1.61
	2	50		0.85	0.95	1.26	1.61
	5	20		1.02	1.00	1.21	1.52
	10	10		1.49	1.11	1.41	1.47
	10	1		0.69	0.78	1.22	1.50
	25	1		0.82	0.69	1.10	1.55
TVP-FRW-FSV	1			1.10	1.41	1.78	1.76
	2			0.94	1.32	1.51	1.54
	5			0.81	1.22	1.37	1.40
	10			0.74	1.16	1.21	1.21
constant with FSV				0.95	1.09	1.65	1.00

Notes: TVP-FBART-FHB refers to the nonparametric model with a factor BART form for the conditional mean and factor heteroBART for the conditional variance. TVP-FRW-FSV denotes the parametric model with TVP and SV processes following factor random walks. The red shaded cell indicates the best model specification, while the gray shaded cell indicates the benchmark (a constant-coefficient VAR with FSV and $Q_q = 12$). The WAIC is evaluated jointly for unemployment (UNRATE), overall consumer price inflation (CPIAUCSL), and the federal funds rate (FEDFUNDS).

is found for models with a moderate number of factors and trees.

The fact that the constant coefficient VAR with FSV can be beaten by TVP models if the number of factors is chosen judiciously indicates that there is parameter change in the VAR coefficients. But the fact that the TVP-FBART-FHB models nearly always exhibit substantially lower WAIC values than the TVP-FRW-FSV models with the same number of factors suggests that the parameter change is better modeled nonparametrically than via random walks.

C.3   Two-year-ahead impulse responses to a business cycle shock

Fig. C.3: Impulse responses over time at the two-year-ahead horizon ($h = 8$).
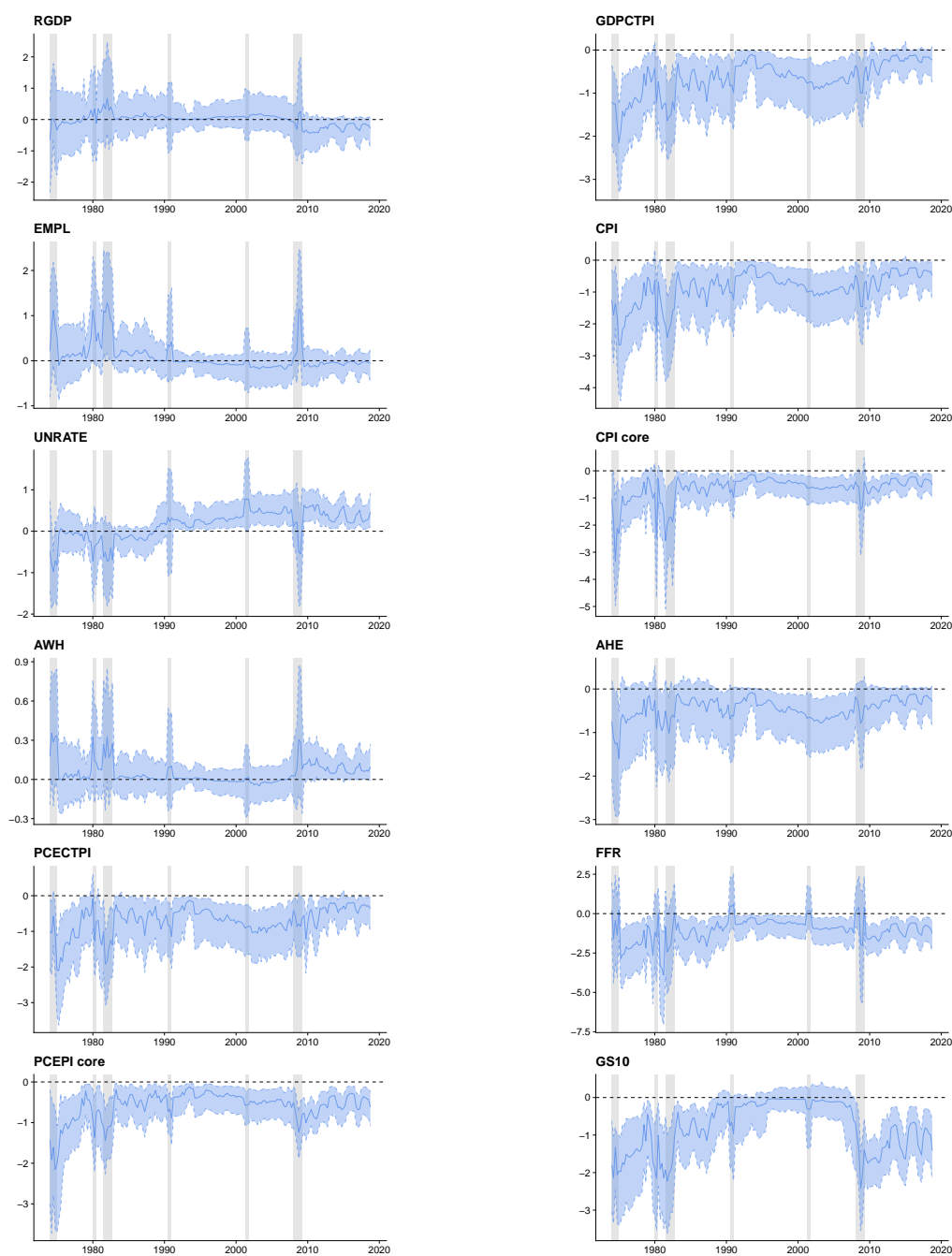
Notes: Time-specific impulse responses to a negative business cycle shock. Blue solid lines denote the posterior median, blue dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the blue shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Panels: endogenous variables. Vertical axis: impulse responses. Front axis: periods (in quarters).

Fig. C.4: Impulse response of prices for different scenarios at the two-year-ahead horizon ($h = 8$).
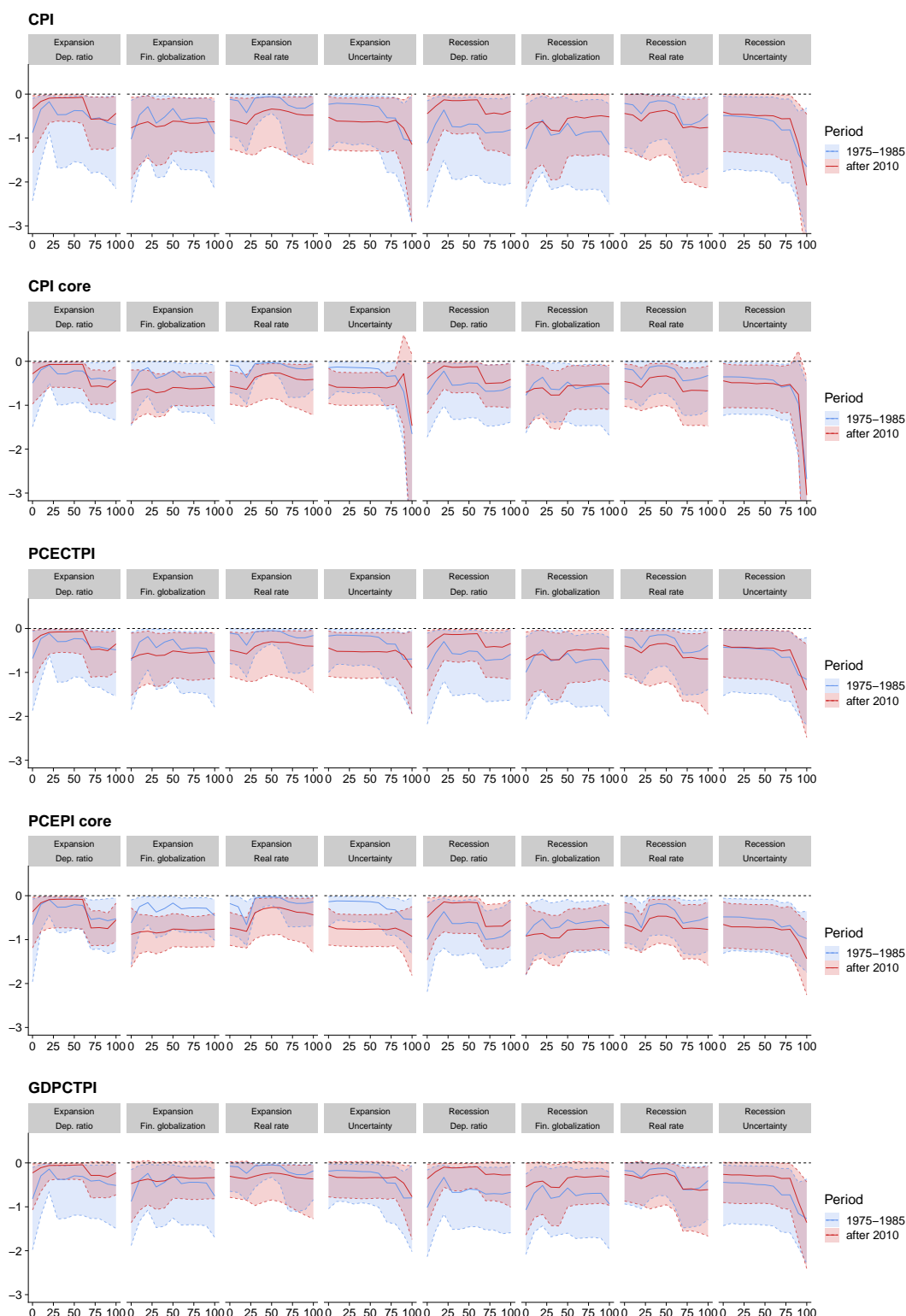
Notes: Impulse responses to a negative business cycle shock by partially varying the effect modifiers. For example, the top-left panel refers to the responses across percentiles of the dependency ratio while assuming an expansion state and setting the remaining effect modifiers (i.e., financial globalization, real interest rate, and uncertainty) either to the mean of the subsample of periods 1975:Q1 to 1984:Q4 (colored in blue) or to the mean of the subsample of periods from 2010:Q1 to 2019:Q4 (colored in red). Colored solid lines denote the posterior median, colored dashed lines the $16^{th}/84^{th}$ posterior percentiles, with the colored shaded areas corresponding to the 68% credible sets, and the black dashed lines mark the zero line. Vertical panels: price indices. Horizontal panels: effect modifiers for an expansion and a recession state. Vertical axis: impulse responses. Front axis: percentiles in % where 0%(100%) denotes the minimum (maximum) value.