Kent Academic Repository

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian nonparametric vector autoregressive models

Maria Kalli [1], Jim E. Griffin [*,1]

School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK

ABSTRACT

Vector autoregressive (VAR) models are the main work-horse models for macroeconomic forecasting, and provide a framework for the analysis of complex dynamics that are present between macroeconomic variables. Whether a classical or a Bayesian approach is adopted, most VAR models are linear with Gaussian innovations. This can limit the model's ability to explain the relationships in macroeconomic series. We propose a nonparametric VAR model that allows for nonlinearity in the conditional mean, heteroscedasticity in the conditional variance, and non-Gaussian innovations. Our approach differs from that of previous studies by modelling the stationary and transition densities using Bayesian nonparametric methods. Our Bayesian nonparametric VAR (BayesNP-VAR) model is applied to US and UK macroeconomic time series, and compared to other Bayesian VAR models. We show that BayesNP-VAR is a flexible model that is able to account for nonlinear relationships as well as heteroscedasticity in the data. In terms of short-run out-of-sample forecasts, we show that BayesNP-VAR predictively outperforms competing models.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Introduced by Sims (1980), vector autoregressive (VAR) models provide a systematic way of capturing the dynamics and interactions of multiple time-series. In its basic form, the L-lag VAR model represents a p-dimensional vector of variables measured at time t, $y_t = (y_{t,1}, \ldots, y_{t,p})'$, as a linear combination of past realisations,

$$y_t = \mu + B_1 y_{t-1} + \cdots + B_L y_{t-L} + e_t \tag{1}$$

where $\{B_l\}_{l=1}^{L}$ are $(p \times p)$-dimensional matrices of unknown coefficients, and $e_t = (e_{1,t}, \ldots, e_{p,t})'$ is a $(p \times 1)$-dimensional innovation vector with distribution $N(0, \Sigma)$. VAR models have emerged as a benchmark for the analysis of dynamic macroeconomic problems. The linear representation of the variables' joint dynamic behaviour facilitates the study of the effects of shocks (such as monetary and fiscal policy shocks) through computation of response functions, and forecast error variance decompositions (see Lucas, 1980; Pagan, 1997; Stock and Watson, 1999; Diebold and Rudebusch, 2001).

Despite their popularity, there have been criticisms of the use of VAR models in macroeconomic analysis. When p is large there is

the risk of overfitting the data which leads to imprecise inference and erratic model forecasts. In addition, the linearity, stationarity, Gaussian innovations and constant conditional mean and variance of these models can be considered unrealistic. For example, empirical evidence suggests that macroeconomic variables may have nonlinear relationships (see Granger and Terasvirta, 1994), the nature of shocks may not be Gaussian (Weise, 1999), and the effects of these shocks may not be linear (see Ravn and Sola (2004) and Matthes and Barnichon (2015) for monetary policy studies, and Sörensen et al. (2001), Auerbach and Gorodnichenko (2013), Baum and Koester (2011), and Gambacorta et al. (2014) for fiscal policy studies).

In the last two decades these criticisms have been addressed by: using adaptations of the parametric model given in Eq. (1) such as regime switching and threshold crossing behaviour, introducing time varying coefficient models with or without stochastic volatility, and, more recently, using nonparametric methods and considering non-Gaussian innovations. Both regime switching and threshold models are motivated by empirical evidence that many macroeconomic time series behave differently during different time periods (for example, in economic downturns and in expansions) which are often called regimes. Both models assume that there are a small number of regimes which can be accurately modelled by different VARs. The mechanism for the change between regimes is the key difference between the two approaches. Hamilton (1989) popularised Markov-switching regression where the change in regimes is driven by latent (unobservable) stochastic

variables, usually with a Markov structure. Literature on these models has subsequently grown, see *e.g.* Hansen (1992), Chib (1996), Chauvet (1998), Kim and Nelson (1999), Kim et al. (2005), and Sims and Zha (2006). Beaudry and Koop (1993), Teräsvirta (1994), Potter (1995), and Pesaran and Potter (1997) popularised vector threshold autoregressive (VTAR) and vector smooth transition autoregressive (VSTAR) models. Unlike Markov-switching regressions, regime changes in a VTAR model occur if some function (often, a linear function) of the observable macroeconomic variables crosses a threshold. Whereas VSTAR uses a weighted average of VAR models where the weighting depends on a continuous, non-linear function of the previous lags. For a comprehensive survey see Hubrich and Teräsvirta (2013). In contrast, time varying vector autoregressions are a class of models in which the system's conditional mean and/or variance are allowed to vary over time. This is achieved by modelling the VAR parameters (coefficients and innovation covariance matrix) with a linear time series model, often a random walk or an AR(1) process. Notable work in this area is presented in Stock and Watson (1996, 2001, 2002), Cogley and Sargent (2001, 2005a), and Primiceri (2005). See Koop and Korobilis (2010) for a recent review of these methods. An alternative approach to modelling the joint dynamic behaviour are nonparametric methods. Härdle et al. (1998) proposed a vector conditional heteroskedastic autoregressive nonlinear model where both the conditional mean and variance are unknown functions of past observations. Hamilton (2001) developed a flexible parametric regression model where the conditional mean has a linear parametric component and a potential nonlinear component represented by an isotropic Gaussian random field. Dahl and Gonzalez-Rivera (2003a, b) extended his model to non-Gaussian random fields, while Jeliazkov (2013) models the conditional mean using a Bayesian hierarchical representation of generalised additive models, where a "smoothness prior" is given to the nonparametric function of the vector of past realisations. The use of non-Gaussian innovations is linked to structural VAR models where the computation of impulse response functions requires identification of the structural errors. Hyvärinen et al. (2010), Moneta et al. (2013), and Lanne et al. (2017) use independent component analysis where they assume mutual independence across the non-Gaussian innovation processes and represent the residuals (obtained when estimating the VAR model) as linear mixtures of these. Lanne and Lütkepohl (2010) model the innovations using a mixture of two Gaussian distributions, whereas Jeliazkov (2013) uses the Student t-distribution.

Geweke and Keane (2007) state that answering interesting questions in economics, from macroeconomic policy to the evaluation of economic welfare, often requires the entire conditional distribution $p(y|x)$. In this paper, we introduce a novel stationary model for multivariate time series where the stationary and transition densities are directly modelled using Bayesian nonparametric methods, which place a prior on an infinite dimensional parameter space and adapt their complexity to the data (see Hjort et al., 2010 for a book length review of Bayesian nonparametric methods). The Bayesian nonparametric approach to density estimation requires a prior on distributions with random smooth densities. We use a Dirichlet process mixture (DPM), the most popular of these priors, which is an infinite mixture model, with the Dirichlet process as the mixing measure. There are several advantages to using Bayesian nonparametric methods. Unlike classical nonparametric methods, there is no need to tune any smoothing parameters. Uncertainty in the unknown density can be expressed through the posterior. The out-of-sample predictive performance of models where the conditional density is estimated using the Bayesian nonparametric approach is superior to other competitive models, see Norets and Pati (2017). Adopting the DPM prior, allows us to construct a mixture of VARs which can be viewed as a multivariate

mixture-of-experts. Introduced by Jacobs et al. (1991) and Jordan and Jacobs (1994), mixture-of-experts models focus on estimating the conditional predictive density $p(y|x)$ for all x where y is univariate (discrete or continuous) and x a high dimensional set of covariates. They are extensions of mixture regression models that allow for covariates in the mixture weights. Geweke and Keane (2007) and Villani et al. (2009) provide extensive analyses when the mixture components are Gaussian, whereas Villani et al. (2012) allow for distributions outside the exponential family to represent the mixture components. In our Bayesian nonparametric mixture of VARs, the mixing weights of the transition density depend on the previous lags allowing different component transition densities to be favoured at different times (for example, in expansionary and contractionary periods) based on lagged observed values. Intuitively, we can view each mixture component ("expert") as a regime with changes of regime determined by the lagged values (through the mixing weights). Our Bayesian nonparametric VAR model allows for nonlinearity in the conditional mean, heteroskedasticity in the conditional variance, and non-Gaussian innovations. We tackle over-parameterisation and the danger of overfitting in two ways, via a prior on the number of mixture components and by modelling the dependence within each component with a prior favouring a simple correlation structure. We find that our approach produces better forecasts (particularly at longer time horizons) when compared to the widely used time varying parameter models with stochastic volatility (TVP-SV-VAR).

The paper is organised as follows: Section 2 introduces the Bayesian non-parametric VAR (BayesNP-VAR) model, describes its construction and considers some of its properties. Section 3 provides an overview of the required Markov chain Monte Carlo (MCMC) method for fitting this model (the full steps of the MCMC sampler are described in Appendix A). Section 4 illustrates the ability of the BayesNP-VAR model to identify regimes and changes in regimes using simulated data, provides an empirical illustration using US macroeconomic time series, and compares the out-of-sample predictive performance of the BayesNP-VAR to the parametric BVAR and TVP-SV-VAR using both US and UK macroeconomic series. Section 5 summarises our findings and conclusions.

2. The Bayesian non-parametric vector autoregressive (BayesNP-VAR) model

We construct a multivariate time series model in which the stationary and transition densities are infinite mixtures. Antoniano-Villalobos and Walker (2016) define such a model for a univariate stationary time series. Their prior has full support for the transition density and stationary density (*i.e.* any transition density and stationary density can be represented arbitrarily well by the prior). We extend their work to multivariate stationary time series and we call our model Bayesian nonparametric VAR (BayesNP-VAR).

The transition densities of the BayesNP-VAR model are derived from a joint distribution for y_t and its L lags y_t^L which is expressed as an infinite mixture. This ensures that the stationary distribution is known and also has the form of an infinite mixture. Specifically, the joint density of y_t and its L lags y_t^L is

$$p(y_t, y_t^L) = \sum_{j=1}^{\infty} w_j \, k(y_t, y_t^L | \theta_j) \qquad (2)$$

where $k(y_t, y_t^L | \theta_j)$ is a $(L+1)p$-dimensional probability density function which does not depend on t and θ_j are the locations of the mixture components with $\theta_j \overset{iid}{\sim} H$. We assume that $k(y_{t-i}, \ldots, y_{t-i-\kappa} | \theta_j)$ for $i = 0, \ldots, L - \kappa$ and $\kappa = 0, \ldots, L - 1$ depends on κ only (which can be achieved by assuming that $k(y_t, y_t^L | \theta_j)$ is the joint distribution of a stationary process) to ensure

that the overall process is stationary. The mixture weights w_j are defined as $w_1 = v_1$, $w_j = v_j \prod_{m<j}(1-v_m)$, and $v_j \overset{iid}{\sim} \text{Be}(1, M)$. Assuming that the locations θ_j are independent of the weights, w_j, the model in Eq. (2) defines a Dirichlet process mixture (Sethuraman, 1994). The distribution of the locations, H, is often referred to as the "base measure", the choice of which determines the likely location of the components. The parameter M controls the relative values of the weights. The expectation of the jth weight is $\text{E}[w_j] = \frac{M^{j-1}}{(M+1)^j}$ and so, as M increases, the average size of the jth weight becomes smaller and the number of components with non-negligible weights becomes larger. Choosing a prior for M is key to controlling the number of components and avoiding overfitting, we discuss this choice later in this section.

The joint density in Eq. (2) leads to a transition density that is also an infinite mixture with the following form:

$$p(y_t|y_t^L) = \frac{p(y_t, y_t^L)}{p(y_t^L)} = \frac{\sum_{j=1}^{\infty} w_j k(y_t, y_t^L|\theta_j)}{\sum_{j=1}^{\infty} w_j k(y_t^L|\theta_j)} \quad (3)$$
$$= \sum_{j=1}^{\infty} \omega_j(y_t^L) k(y_t|y_t^L, \theta_j)$$

where $k(y_t|y_t^L, \theta_j)$ is the transition density of the jth component and $\omega_j(y_t^L) = \frac{w_j k(y_t^L|\theta_j)}{\sum_{k=1}^{\infty} w_k k(y_t^L|\theta_k)}$ is the weight of the jth component which depends on previous lags, the key feature of our model. We can therefore refer to the transition density as a multivariate mixture of experts. Mixtures of experts are extensions of smooth regression models and popular within the machine learning community. They are used in regression to estimate the conditional density $p(y|x)$ of a univariate y for all values of a (often, high-dimensional) covariate x, using mixtures where the component weights depend on a x, see Jacobs et al. (1991), Jordan and Jacobs (1994), Geweke and Keane (2007) and Villani et al. (2012). The weights of the transition density in Eq. (3) depend on the observed lagged values. This feature allows different component transition densities to be favoured in different periods. For example, expansionary and recessionary periods could have different transition densities. In our multivariate time series model, each component ("expert") can be viewed as a regime with changes of regime determined by the observed lagged values.

To complete the BayesNP-VAR we need to choose $k(y_t, y_t^L|\theta_j)$. Firstly, we find it easier to write $y_t = \mu + S\epsilon_t$ where μ is a p-dimensional location vector and $S = \text{diag}(s_1, \ldots, s_p)$ is a scaling matrix. Then $k(y_t, y_t^L|\theta_j) = \left(\prod_{j=1}^{p} s_j\right)^{-(L+1)} k_\epsilon(S^{-1}(y_t - \mu), (S^{-1} \otimes I_L)(y_t - 1_L\mu))$ where k_ϵ is the joint density of ϵ_t, ϵ_t^L. We propose a model which has a structure similar to a factor model and divides the variation of the data into a part which describes the dependence between variables and a part which is idiosyncratic to each variable. The assumed form is $k_\epsilon(\epsilon_t, \epsilon_t^L) = \text{N}(0_{(L+1)p}, B + Q)$ where $0_{(L+1)p}$ is a $(L+1)p$-dimensional vector of 0's, B describes the dependence between variables over time and Q describes the idiosyncratic variation over time. The form of B is

$$B = \sum_{z=1}^{q} P_z \otimes (\Lambda_{.,z} \Lambda_{.,z}^T)$$

where

$$P_z = \begin{pmatrix} 1 & \rho_z & \ldots & \rho_z^L \\ \rho_z & 1 & \ldots & \rho_z^{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_z^L & \rho_z^{L-1} & \ldots & 1 \end{pmatrix},$$

$-1 < \rho_z < 1$ for $z = 1, \ldots, q$ and Λ is a $(p \times q)$-dimensional matrix of loadings. The matrix Q is

$$Q = \begin{pmatrix} \Sigma_0 & \Sigma_1 & \Sigma_2 & \ldots & \Sigma_L \\ \Sigma_1 & \Sigma_0 & \Sigma_1 & \ldots & \Sigma_{L-1} \\ \Sigma_2 & \Sigma_1 & \Sigma_0 & \ldots & \Sigma_{L-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_L & \Sigma_{L-1} & \Sigma_{L-2} & \ldots & \Sigma_0 \end{pmatrix},$$

where $\Sigma_l = \text{diag}\left(\frac{\rho_1^{\star l}}{\xi_1}, \frac{\rho_2^{\star l}}{\xi_2}, \ldots, \frac{\rho_p^{\star l}}{\xi_p}\right)$ for $l = 1, \ldots, L$, $-1 < \rho_k^\star < 1$ and $\xi_k^{-1} \sim \text{Ga}(v/2, v/2)$ for $k = 1, \ldots, p$. This leads to a suitable choice of $k(y_t, y_t^L)$ for the BayesNP-VAR model to be stationary. The marginal distribution is $k_\epsilon(\epsilon_t) = \text{N}\left(0, \Lambda\Lambda^T + \text{diag}(\xi_1^{-1}, \xi_2^{-1}, \ldots, \xi_p^{-1})\right)$ and so the marginal distribution of y_t is $\text{N}\left(\mu, S\Lambda\Lambda^T S^T + \text{diag}(s_1^2\xi_1^{-1}, s_2^2\xi_2^{-1}, \ldots, s_p^2\xi_p^{-1})\right)$.

The component-specific parameters in the mixture model are $\theta = (\mu, S, \Lambda, \rho, \rho^\star, \phi, \delta)$ where $\rho = (\rho_1, \ldots, \rho_q)$, $\rho^\star = (\rho_1^\star, \ldots, \rho_p^\star)$ and $\delta = (\delta_1, \ldots, \delta_q)$. We assume that $\mu, S, \Lambda, \rho, \rho^\star, \phi,$ and δ are *a priori* independent with distribution H which has density,

$$h(\mu, S, \Lambda, \rho, \rho^\star, \phi, \delta) = h_\mu(\mu) \times h_S(S) \times h_\Lambda(\Lambda)$$
$$\times h_{\rho^\star}(\rho^\star) \times h_\rho(\rho) \times h_\phi(\phi) \times h_\delta(\delta).$$

The parameters μ and S are given informative prior densities to avoid the mixture model placing mass in areas which are not plausible. We choose $h_\mu(\mu) = \text{N}(\mu|\mu_0, \Sigma_0)$. Both parameters can be chosen with prior information but we use the data dependent choices $\mu_0 = \bar{y}_t$, and $\Sigma_0 = 1.5^2 \widehat{\Sigma}$, where $\bar{y}_t$, and $\widehat{\Sigma}$ are sample mean and covariance matrix of y_t in the empirical examples in this paper. This choice leads to a prior for μ which is slightly overdispersed relative to the distribution of the data, and so the components are located in regions within or close to the data. For the $(p \times p)$-dimensional scaling matrix, $S = \text{diag}(s_1, s_2, \ldots, s_p)$, we choose the hierarchical prior

$$s_i^{-1} \overset{ind.}{\sim} \text{Ga}(a_s, \zeta_i(a_s - 1)), \qquad \zeta_i \overset{iid}{\sim} \text{Ga}(1, 5).$$

The hyperparameter ζ_i is shared by all components and is an estimate of the overall scale of the ith variable. This hierarchical structure allows different components to have similar scales, the s_i's, for each variable.

We use the multiplicative gamma process shrinkage prior of Bhattacharya and Dunson (2011) for Λ. This allows the complexity of B to adapt to the data. Under this prior, $\Lambda_{i,z} \overset{iid}{\sim} \text{N}(0, \phi_{i,z}^{-1}\tau_z^{-1})$, for $i = 1, \ldots, p$ and $z = 1, \ldots, q$ where $\phi_{i,z} \sim \text{Ga}(v/2, v/2)$ and $\tau_z = \prod_{i=1}^{z} \delta_i$ with $\delta_1 \sim \text{Ga}(1, 1)$ and $\delta_z \sim \text{Ga}(3, 1)$ for $z \geq 2$. The δ_z's are independent, and the τ_z's are viewed as the global shrinkage parameters of the columns, while $\phi_{i,z}$'s are the local shrinkage parameters for the zth column. As value of δ_z increases, so does the value of τ_z favouring smaller values of $\Lambda_{i,z}$.

The parameters ρ and $\rho^\star$ control dependence across time in both B and Q. We choose independent uniform priors on the range that implies stationarity and positive autocorrelation to give $h_\rho(\rho) = \prod_{z=1}^{q} \text{U}(\rho_z|0, 1)$ and $h_{\rho^\star}(\rho^\star) = \prod_{i=1}^{p} \text{U}(\rho_i^\star|0, 1)$ where $\text{U}(\rho|0, 1)$ represents the density of a uniform distribution between 0 and 1. For the prior of M, the parameter controlling the number of components, we choose the standard exponential distribution. We find that this choice strikes a balance between having too many and/or too few components.

3. Inference in Bayesian NP-VAR

The likelihood function can be derived from the transition in Eq. (3) to be defined as

$$\prod_{t=L+1}^{T} p(y_t|y_t^L).$$

Bayesian inference is complicated by the infinite sum in both the numerator and denominator which precludes the direct use of Markov chain Monte Carlo methods. Antoniano-Villalobos and Walker (2016) describe a Gibbs sampler for their univariate model but truncate the centring distribution for the stationary variance of each component away from zero. To avoid this truncation, we use an adaptive truncation method introduced by Griffin (2016) which adaptively truncates the infinite sum in the numerator and denominator and tends to avoid large truncation errors in the posterior. We define a truncation of the infinite model in Eq. (2) with K mixture components which leads to a truncated transition density which has the form

$$p_K(y_t|y_t^L) = \frac{\sum_{j=1}^{K} w_j k\left(y_t, y_t^L|\theta_j\right)}{\sum_{j=1}^{K} w_j k\left(y_t^L|\theta_j\right)} \tag{4}$$

where $w_j = V_j \prod_{m<j}(1 - V_m)$ and $V_j \overset{iid}{\sim} Be(1, M)$ and $\theta_j \overset{iid}{\sim} H$. This finite mixture model defines a sequence of posteriors of the form

$$\pi_K(\theta_{1:K}, \eta_{1:K}|y) \propto p_K(\theta_{1:K}, \eta_{1:K}) \prod_{t=L+1}^{T} p_K(y_t|y_t^L)$$

where $\eta_{1:K} = (V_{1:K}, \zeta, M)$. The adaptive truncation method of Griffin (2016) uses an MCMC algorithm to sample from the posterior, $\pi_{K_0}(\theta_{1:K_0}, \eta_{1:K_0}|y)$, for a user-defined starting value, K_0, and then uses a sequential Monte Carlo method to sample from the sequence of posterior distributions $\pi_{K_0+1}(\theta_{1:(K_0+1)}, \eta_{1:(K_0+1)}|y)$, $\pi_{K_0+2}(\theta_{1:(K_0+2)}, \eta_{1:(K_0+2)}|y), \ldots, \pi_D(\theta_{1:D}, \eta_{1:D}|y)$ where D is chosen to avoid large truncation errors. The adaptive truncation scheme follows the algorithm below.

1. Simulate a sample of size N using the MCMC sampler from $\pi_{K_0}(\theta_{1:K_0}, \eta_{1:K_0}|y)$ which will be denoted $\left(\theta_{1:K_0}^{(1)}, \eta_{1:K_0}^{(1)}\right), \ldots,$ $\left(\theta_{1:K_0}^{(N)}, \eta_{1:K_0}^{(N)}\right)$ and set $K = K_0 + 1$.

2. Simulate $\left(\theta_K^{(i)}, \phi_K^{(i)}, V_K^{(i)}\right)$ from their prior distribution for $i = 1, \ldots, N$.

3. Evaluate

$$\psi_i = \frac{\pi_K\left(\theta_{1:K}^{(i)}, \eta_{1:K}^{(i)}|y\right)}{\pi_{K-1}\left(\theta_{1:(K-1)}^{(i)}, \eta_{1:(K-1)}^{(i)}|y\right)}, \qquad i = 1, \ldots, N$$

4. Evaluate

$$\text{ESS}_K = \frac{\left(\sum_{i=1}^{N} \psi_i\right)^2}{\sum_{i=1}^{N} \psi_i^2}.$$

5. If $\text{ESS}_K < cN$ (we use $c = 0.5$) then generate N values where $\left(\theta_K^{(i)}, \eta_K^{(i)}\right)$ is sampled with probability proportional to ψ_i. Set $\psi_i = 1$ for $i = 1, \ldots, N$ and run one iteration of the MCMC sampler updating $\left(\theta_{1:K}^{(i)}, \eta_{1:K}^{(i)}\right)$ from $\pi_K(\theta_{1:K}, \eta_{1:K}|y)$ for $i = 1, \ldots, N$.

6. Let $\Delta_K = |\text{ESS}_K - \text{ESS}_{K-1}|$. If $\Delta_K \leq \varepsilon$, $\Delta_{K-1} \leq \varepsilon$ and $\Delta_{K-2} \leq \varepsilon$ terminate (we choose ε equal to 0.001N). Otherwise, set $K = K + 1$, and return to step 2.

Full details of the MCMC algorithm are provided in Appendix A. The MCMC algorithm uses two types of adaptive Metropolis–Hastings algorithm which are briefly reviewed here. The first is the adaptive random walk Metropolis–Hastings algorithm (Atchadé and Rosenthal, 2005) with a normal proposal whose variance is updated during the running of the chain. Suppose that σ_t^2 is the proposal variance used at iteration t, then the proposal variance at time $t + 1$ is $\sigma_{t+1}^2 = \sigma_t^2 + t^{-0.6}(\alpha_t - 0.234)$ where α_t is the acceptance probability in the Metropolis–Hastings algorithm at the

tth iteration. Atchadé and Rosenthal (2005) show that this algorithm is ergodic. The second algorithm is Adaptive scaling within the Adaptive Metropolis–Hastings (ASWAM) algorithm (Andrieu and Moulines, 2006; Atchadé and Fort, 2010). This is suitable for updating multiple parameters jointly. Suppose we wish to sample a parameter, λ, then the proposed value λ' at the tth iteration is

$$\lambda' \sim N(\lambda, s_t^2 \Sigma_t)$$

where s_t is a scalar and Σ_t is the sample covariance matrix of the first $t - 1$ sampled values of λ sampled from the posterior. The scale s_t is updated using the recursion $s_{t+1} = s_t + t^{-0.6}(\alpha_t - 0.234)$ where, again, α_t is the acceptance probability of the Metropolis–Hastings algorithm at the tth iteration.

4. Illustrations

In this section we apply the BayesNP-VAR model to both simulated and empirical data. Our aim is to demonstrate that our model identifies economic regimes where shocks are transmitted in different ways, that it clearly indicates changes between regimes, and provide evidence of the model's good out-of-sample predictive performance.

To provide a point estimate of our mixture model, we approximate the posterior mode by selecting the MCMC sample which has the highest posterior density value. We refer to this as the posterior modal sample. This allows us to illustrate the model's ability to correctly identify regimes by producing time plots showing how the weights in the mixture model in Eq. (3) change over time and to highlight the component/regime which is favoured in a particular time period (by finding the component with the highest posterior weight). Since the transition density within each component is a VAR model, we also plot impulse response functions (IRFs) to a unit shock in a chosen variable for each component to understand the different ways that shocks are transmitted under different components/regimes. The usual methodology for generating IRFs is used which involves a polynomial function of the estimated VAR parameters.

The posterior mode is useful for presenting the inference from our model but it ignores posterior uncertainty. To provide IRFs which include posterior uncertainty, we also produce the IRFs suggested by Koop et al. (1996) which we will refer to as Generalised IRFs (GIRFs). These are used to study the effect of a shock of size υ at time t on the state of the system at time $t + n$, given that there are no other shocks to the system after time t. They are defined as follows:

$$\begin{aligned} &GI_y(n, \upsilon, Y_{t-1}) \\ &= E[Y_{t+n}|e_t = \upsilon, e_{t+1} = 0, \ldots, e_{t+n} = 0, Y_{t-1}] \\ &\quad - E[Y_{t+n}|e_t = 0, e_{t+1} = 0, \ldots, e_{t+n} = 0, Y_{t-1}] \end{aligned} \tag{5}$$

for $n = 1, 2, 3, \ldots$, where $e_t, e_{t+1}, \ldots, e_{t+n}$ represent the arbitrary shocks at times $t, t + 1, \ldots, t + n$ and the expectations are taken conditional on the parameter values. This allows us to look at posterior distributions of $GI_y(n, \upsilon, Y_{t-1})$ to assess uncertainty in their estimation.

For the out-of-sample predictive performance illustration we calculate the log-predictive score (LPS) for all the variables (using their joint predictive distribution) and for each variable separately (using their marginal predictive distribution). We also calculate the root mean squared error (RMSE) for each variable. The joint LPS is given by

$$-\sum_{i=s}^{T-h} \log p(y_{i+h}|y_1, \ldots, y_i) \tag{6}$$

Table 1
US data; Source: FRED.

Name	Description	Growth rates	Levels
UNRATE	Unemployment rate	None	None
PCEPI	Personal Consumption Expenditure Index: 2009 = 100	$1200\,ln(\frac{y_t}{y_{t-1}})$	$1200\,ln(y_t)$
PAYEMS	Total non-farm payroll, thousands of persons	$1200\,ln(\frac{y_t}{y_{t-1}})$	$1200\,ln(y_t)$
FEDFUNDS	Federal funds rate	None	None
INDPRO	Industrial production index: 2012 = 100	$1200\,ln(\frac{y_t}{y_{t-1}})$	$1200\,ln(y_t)$
LTIR	Long term interest rate	None	None

Table 2
UK data; Source: OECD.

Name	Description	Growth rates	Levels
UNRATE	Unemployment rate	None	None
CPI	Consumer price Index: 2010 = 100	$1200\,ln(\frac{y_t}{y_{t-1}})$	$1200\,ln(y_t)$
STIR	Short term interest rate	None	None
INDPRO	Industrial production index: 2010 = 100	$1200\,ln(\frac{y_t}{y_{t-1}})$	$1200\,ln(y_t)$
LTIR	Long term interest rate	None	None

where T is the size of the time series, s is the time from where the prediction starts, and h is the predictive horizon. Similarly, the marginal LPS for the jth variable is given by

$$-\sum_{i=s}^{T-h} \log p(y_{i+h,j}|y_1, \ldots, y_i). \tag{7}$$

The RMSE of the jth variable is given by

$$\sqrt{\frac{1}{T-h-s+1}\sum_{i=s}^{T-h}(y_{i+h,j} - E(y_{i+h,j}|y_1, \ldots, y_i))^2}. \tag{8}$$

For both measures, smaller scores indicate better predictive performance. We looked at $h = 1, 2,$ and 4 months. We compare the out-of-sample predictive performance of our BayesNP-VAR model with other Bayesian VAR specifications: the stationary BVAR model with the independent Normal-Wishart prior (with either one, two, three or four lags), and the non-stationary non-linear TVP-VAR model with stochastic volatility (TVP-SV-VAR) of Primiceri (2005) (with one lag). The BayesNP-VAR was chosen to have one lag.

4.1. Simulated data

The following simulated example illustrates the ability of the BayesNP-VAR to correctly identify regimes and the timing of regime switches. We generated data from a threshold VAR(2) model with $p = 3$ variables and 500 time points. The data had two regimes and followed the VAR in Regime 1 if $y_{t-1,1} > 0$ and Regime 2 otherwise. The two regimes were

- Regime 1

$$y_t = \begin{pmatrix} 1.8 \\ 0.52 \\ 0.29 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.15 & 0 \\ 0.20 & 0.34 & 0 \\ 0.03 & 0.05 & 0.24 \end{pmatrix} y_{t-1}$$
$$+ \begin{pmatrix} 0.15 & 0.20 & 0.80 \\ 0.14 & -0.18 & 0.30 \\ 0.07 & -0.03 & 0.14 \end{pmatrix} y_{t-2} + e_t$$

with covariance matrix,

$$\Sigma_1 = \begin{pmatrix} 0.28 & 0.03 & 0.07 \\ 0.03 & 0.29 & 0.14 \\ 0.07 & 0.14 & 0.36 \end{pmatrix}$$

- Regime 2

$$y_t = \begin{pmatrix} -1.8 \\ 0.32 \\ 0.12 \end{pmatrix} + \begin{pmatrix} 0.6 & -0.05 & 0.2 \\ 0.20 & 0.09 & 0 \\ 0.05 & 0 & 0.42 \end{pmatrix} y_{t-1}$$
$$+ \begin{pmatrix} 0.21 & -0.10 & 0.05 \\ 0.07 & 0.32 & 0 \\ 0.06 & -0.02 & 0.45 \end{pmatrix} y_{t-2} + e_t$$

with covariance matrix,

$$\Sigma_2 = \begin{pmatrix} 0.54 & 0.06 & 0.02 \\ 0.06 & 0.46 & 0.24 \\ 0.02 & 0.024 & 0.56 \end{pmatrix}.$$

Fig. 1 displays plots of the three simulated series and some results for the posterior modal sample which identifies two components with non-negligible mixing weights. The top row of each plot shows the time series for each variable and periods for which a component has the highest weight. The second row of plots shows the weight for that regime (and is the same for all variables). The threshold is indicated by a dashed line and was set at $y_{t-1,1} = 0$. We can clearly see that the estimated regimes change correctly as the value of $y_{t-1,1}$ changes. The second row displays the weight of the regime and we can see that the regimes are correctly identified with probability close to 1.

We also simulated a VAR(2) model with $p = 3$ variables and 500 time points with the specification given in Appendix B. We do not display the results here but our model is able to correctly identify that there is only one regime. Both simulated threshold VAR(2) and VAR(2) data are used in Section 4.3 as part of our out-of-sample forecasting exercise.

4.2. Empirical examples

4.2.1. Data sets

We constructed two macroeconomic data sets, one for the US and one for the UK, based on the series described and transformations carried out in Carriero et al. (2015). Both data sets include seasonally adjusted monthly time series, the sample period for the US is from 1st January 1959 to 1st August 2016, and for the UK from 1st January 1978 to 1st February 2015. The US series were collected from the Federal Reserve Bank of St Louis (FRED) and UK series from the Organisation for Economic Cooperation and Development (OECD). Details of the variables together with the transformations used are displayed in Table 1 for the US and Table 2 for the UK respectively. We include the series in both growth rates and levels.

Component/Regime 1

Component 2/Regime 2
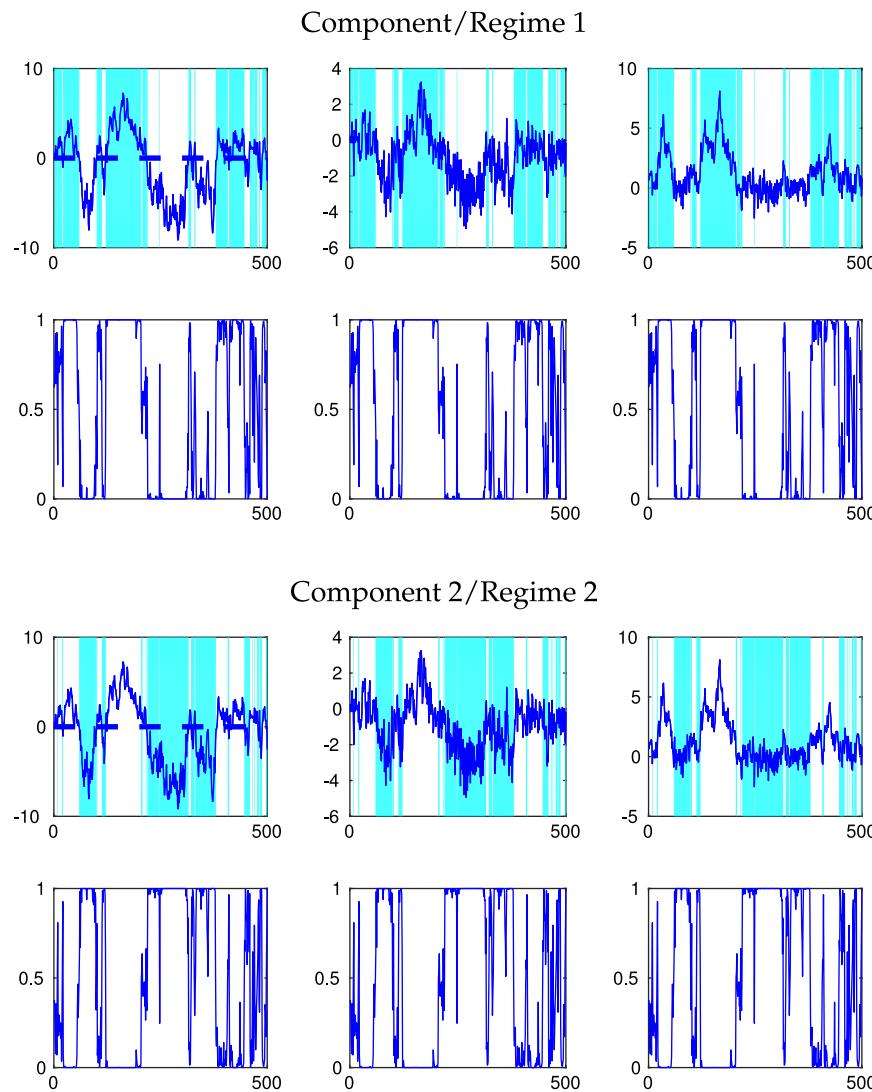
Fig. 1. Plots identifying the two regimes of the simulated threshold VAR(2) series. Top panel displays the three series (left-series 1, middle-series 2 and right-series 3) and highlights (in cyan) the regime. The dashed line at zero in the series 1 plot is the threshold. Bottom panel displays the non-negligible weight of that regime.

For the illustrations in Sections 4.2.2 and 4.2.3 we used the US data, in growth rates. For the out-of-sample predictive exercise we have used both the US and UK data in growth rates and levels.

### 4.2.2. Component/Regime identification

Applying our BayesNP-VAR model to the US data in growth rates, we identify eight mixture components, and here we present the six which had the largest non-negligible weights. Figs. 2–4 display results for the six distinct components. The first two rows show the time series for each variable with the component/regime with the largest mixture weight and the weight for that component/regime is plotted over time (in the bottom row). The first regime, which we named, stable inflation (PCEPI-growth) and output (INDPROD) growth, covers periods of sustained growth and includes *"The Great Moderation"* of the mid-1980s to the mid-2000s, a period when volatility of business cycle fluctuations was low. The second component identifies periods after recent stock market crashes, caused by the burst of the 'Dot.com' bubble and the collapse of the US housing market respectively. These are periods of very low federal funds rates and long term interest rates with fairly stable growth and declining unemployment rate. These are in line with the National Bureau of Economic Research (NBER) peak to through periods of March 2001 to November 2001, and

November 2007 to June 2009 respectively. The *"Golden Era"* of US capitalism, the period up to the early 1970s, is captured by the third component, whereas the fourth component captures the period of mid 1981 to 1984 of the *"Volcker disinflation"*. The last two components identify two of the worst recessions in the US where output growth was in decline and unemployment rapidly increasing. The *"US housing crisis"* of 2007 is captured by the fifth component, and the *"Oil Shock"* of the early 1970s is captured by the sixth component. The difference between the two is that the latter was characterised by high inflation.

4.2.3. Impulse response functions

Each component of our BayesNP-VAR model follows a VAR and so we can produce component/regime dependent IRFs as polynomial functions of the estimated VAR parameters. All IRFs look at 60 months ahead. Figs. 5 and 6 display these IRFs for each component using a different colour and pattern for each. The colour and pattern scheme is: blue dash inverted triangle is for the first component, red dash for the second, green dot dash for the third, yellow solid for the fourth, cyan circle dash for the fifth and pink plus dash for the sixth component. Recall that periods of stable inflation and output growth are in the first component, periods after the stock market crashes of 2001 and 2007 are in

Component 1: Stable inflation and output growth
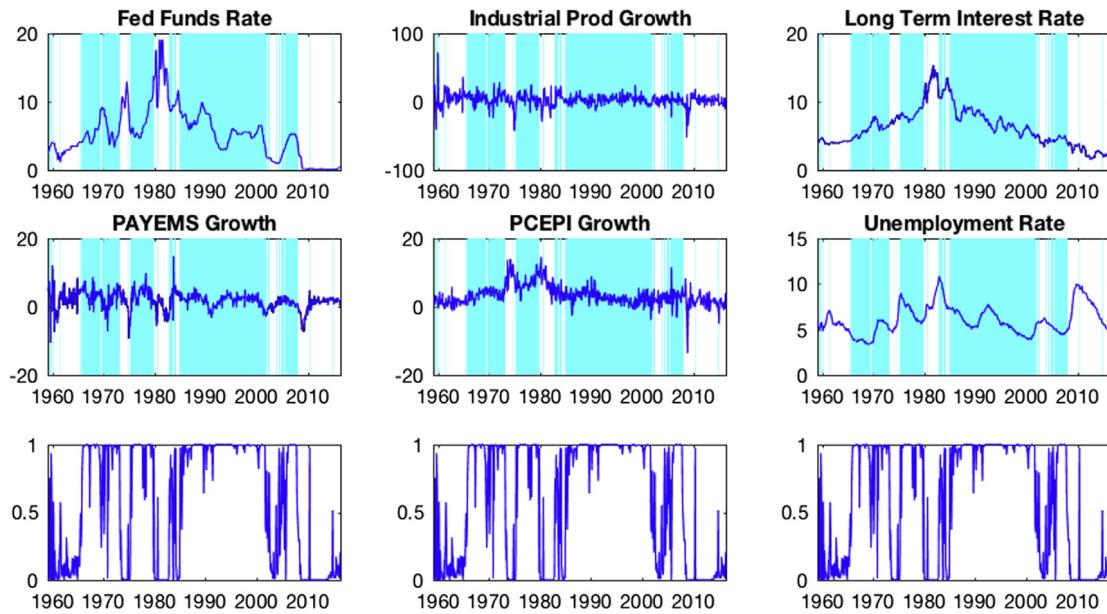
Component 2: After recent stock market crashes
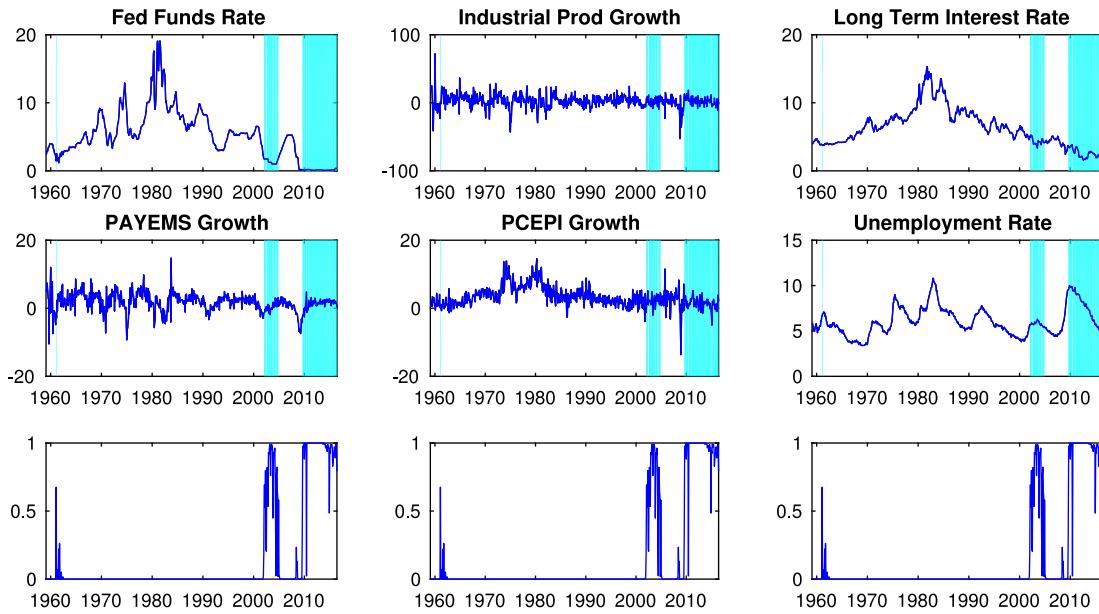
Fig. 2. Plots identifying the first and second components of the US data in growth rates. First two rows of each set of nine plots display the time series highlighting (in cyan) the component/regime, and the third row displays the non-negligible weight of the respective regime.

the second, the "*Golden Era*" periods in the third, the "*Volcker disinflation*" periods in the fourth, the periods of the US housing crisis in the fifth and the 1970s "*Oil shock*" is the sixth component. We centre our IRFs around the study of monetary policy. Fig. 5 displays the IRFs of inflation panel (a), output growth panel (b), and unemployment rate panel (c) to a 1% increase in the federal funds rate. It is clear that the IRFs are regime dependent with the transmission of the monetary policy shock differing between periods of expansion, stability and recession. Inflation eventually goes down and output growth declines in the two crisis periods of 1973 and 2007, whereas the monetary policy shock has small effect on inflation during the periods after the stock market crashes

of 2001 and 2007, Volcker's chairmanship and the "*Golden Era*". During periods of relative stability inflation and output growth increase marginally before remaining at a constant level. The un-employment rate response steadily increases for the "*Golden Era*" and 2007 crisis, is relatively flat for the "*Oil shock*" and "*Volcker disinflation*" regimes, and declines for the component identifying the periods after the stock market crashes of 2001 and 2007, and the component identifying periods of relative economic stability. Fig. 6 displays the IRFs of the federal funds rate to a 1% increase in inflation panel (a) and to a 1% increase in unemployment rate panel (b). During the Volcker disinflation regime, the rate reacts quickly to both an inflationary and an unemployment shock, though the

Component 3: Golden era
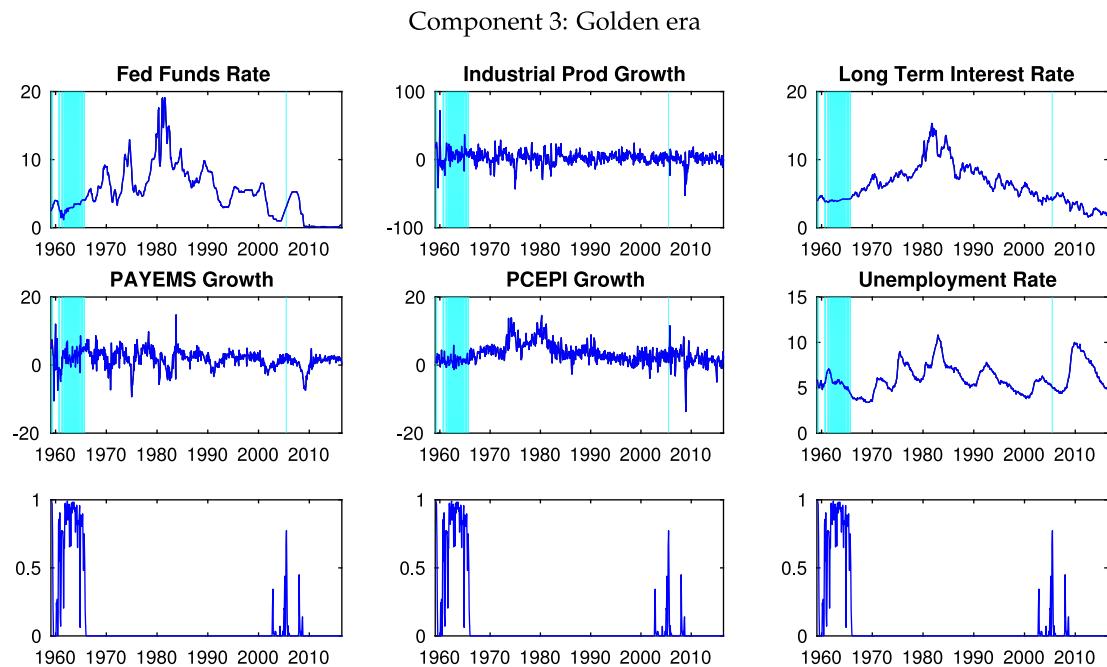
Component 4: Volcker disinflation
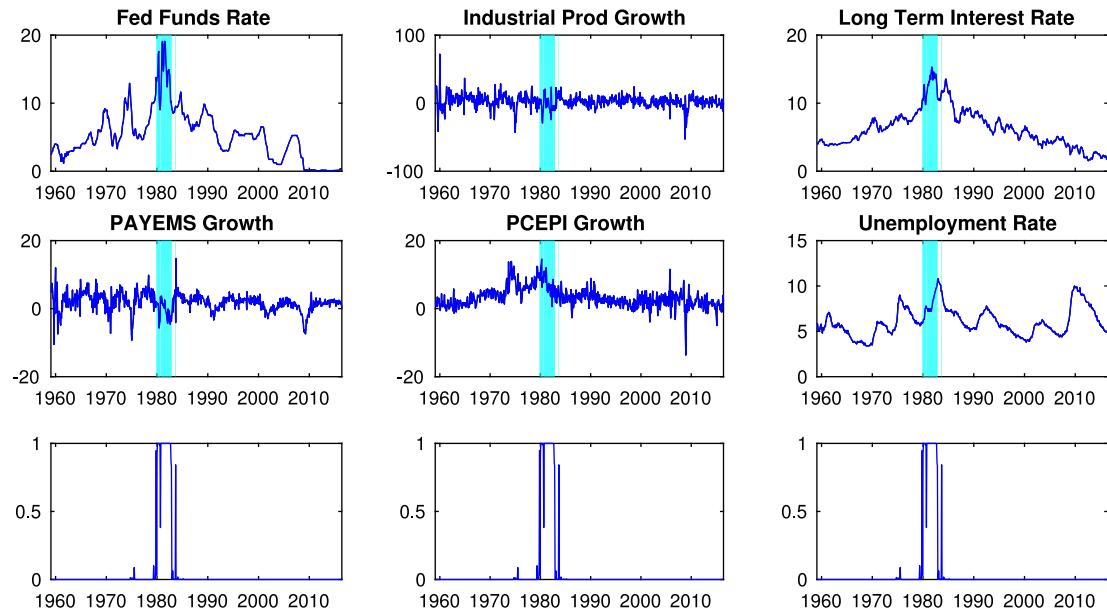
Fig. 3. Plots identifying the third and fourth components of the US data in growth rates. First two rows of each set of nine plots display the time series highlighting (in cyan) the component/regime, and the third row displays the non-negligible weight of the respective regime.

length of time before stabilising is longer for the latter (20 rather than 10 months). A similar pattern can be seen in the *"Golden Era"* regime, however the response to the unemployment shock is more prolonged with a sharp decline for the first 5 months. The rate response to the inflationary shock is not as extreme. When it comes to the other four regimes the inflationary shock has little impact, but the unemployment shock leads to a steady decline in the federal funds rate in the period of stable inflation and output growth, and to a marginal decline for the first 5 months before levelling off for the 2007 US housing crisis regime.

Figs. 7 and 8 display the GIRFs (see Eq. (5)) with their 95% credible intervals for inflation and unemployment rate after a 1% increase in the federal funds rate at two different dates. The dates chosen for comparison are June 1981 and December 2007, the former because it is representative of Paul Volcker's chairmanship of the Fed and the latter because it is the beginning of the US housing crisis. Fig. 7 displays the response of inflation panel (a) and unemployment panel (b), over a 60 month period, to a 1% permanent increase in the Federal funds rate occurring in June 1981. There is no evidence of an effect. Fig. 8 draws the same graphs for a 1% increase in the Federal funds rate occurring in December 2007. There is clear evidence of an increase in inflation with the median response being a 0.5% increase. There is evidence of a negative effect on unemployment with the median effect close to zero, but the lower credible interval shows that a 1% drop in unemployment is plausible.

Component 5: The 2007 housing market crash
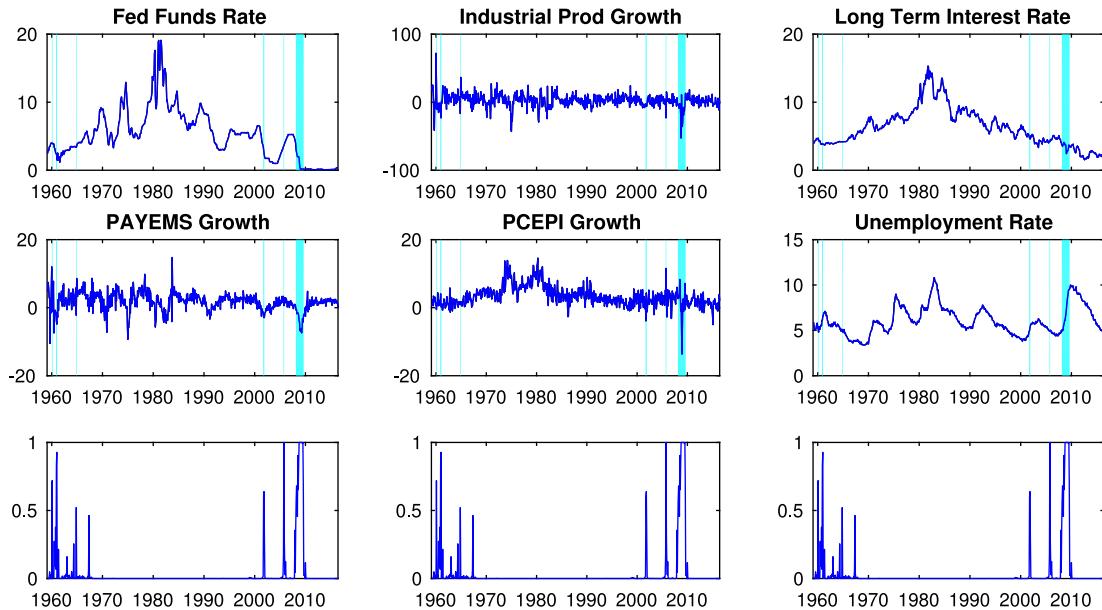
## Component 6: The 1973 oil crisis
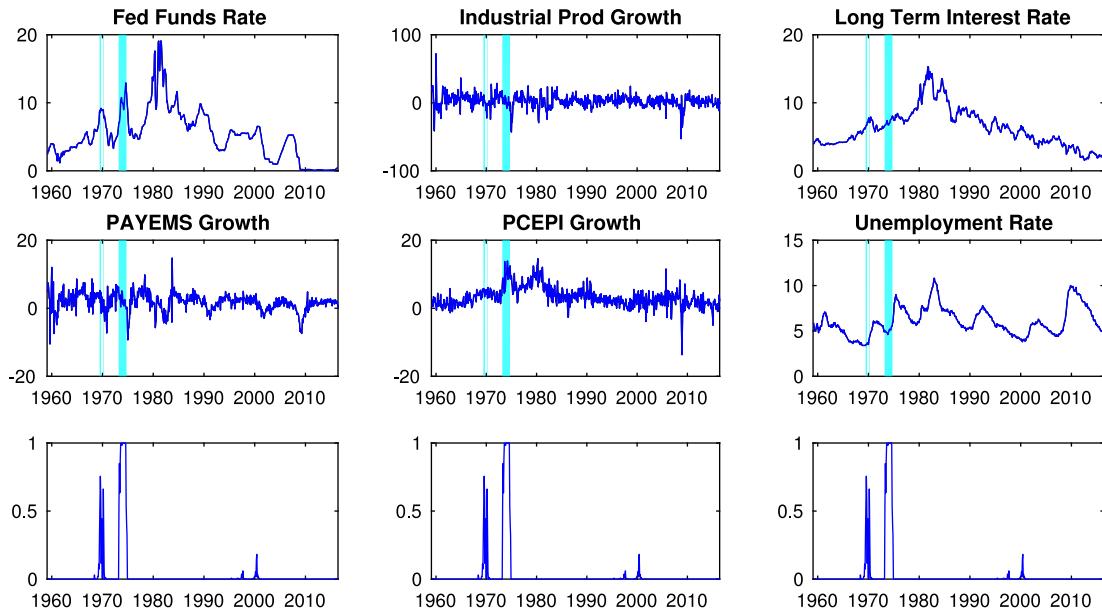
Fig. 4. Plots identifying the fifth and sixth components of the US data in growth rates. First two rows of each set of nine plots display the time series highlighting (in cyan) the component/regime, and the third row displays the non-negligible weight of the respective regime.

4.3. Out-of-sample predictive performance

We compare the out-of-sample predictive performance of the BayesNP-VAR model with other Bayesian VAR specifications: the stationary BVAR model with the independent Normal-Wishart prior (with one, two, three or four lags), and the non-stationary, non-linear TVP-VAR model with stochastic volatility (TVP-SV-VAR) of Primiceri (2005) (with one lag). Our comparison metrics are the log-predictive scores described in Eqs. (6) and (7) and the root mean squared error in Eq. (8). We used two simulated (from threshold VAR(2) and VAR(2) models) and four real data sets (US and UK data in growth rates and levels). We consider three predictive horizons: 1 month, 2 months and 4 months, and look

at 48 months out-of-sample. This means that for the US data the prediction starts on 1st September 2012, and for the UK data on 1st March 2011.

Tables 3 and 4 display the log-predictive scores and the RMSEs for the simulated VAR(2) data and simulated threshold VAR(2) data respectively. The BayesNP-VAR(1) outperforms the TVP-SV-VAR(1) for all predictive horizons both for all variables jointly and each variable marginally, in both the threshold VAR(2) and VAR(2) simulated data. In the VAR(2) data, the overall log-predictive scores and the RMSEs for the BVAR(1) and the BayesNP-VAR(1) are comparable at all horizons. In the threshold VAR(2) data, the BayesNP-VAR(1) has the lowest overall log-predictive scores. It also
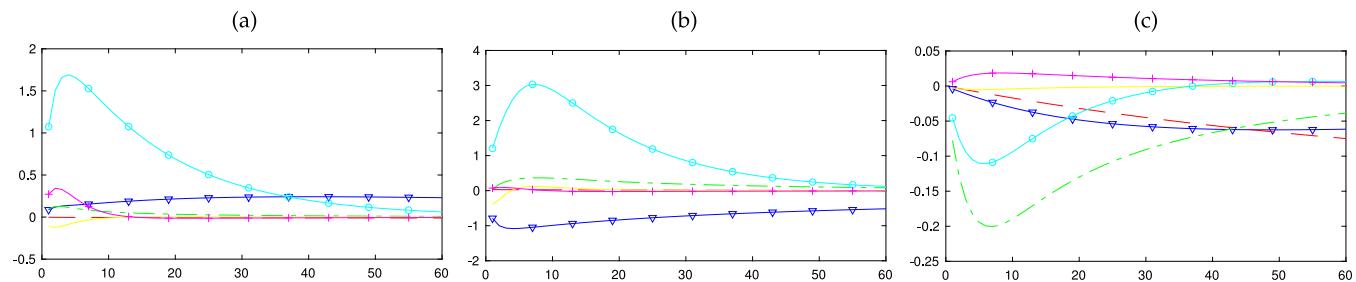
Fig. 5. IRFs to a 1% increase in federal funds rate. (a) Inflation response, (b) Industrial production growth response, and (c) Unemployment response. Blue dash inverted triangle—component 1, Red dash—component 2, Green dot dash—component 3, Yellow solid—component 4, Cyan circle dash—component 5, and Pink plus dash—component 6.
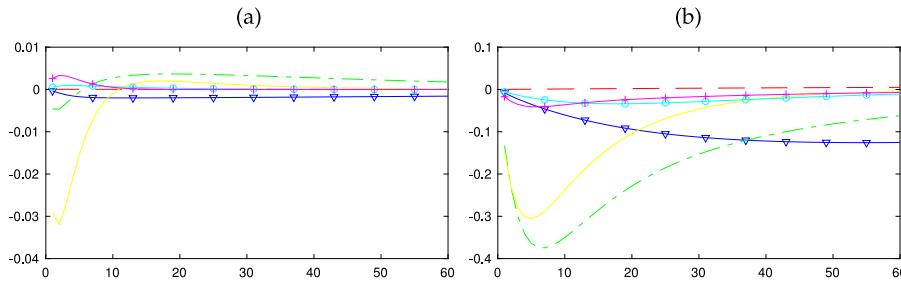
Fig. 6. (a) Federal funds rate response to a 1% increase in inflation, (b) Federal funds rates response to a 1% increase in unemployment rate. Blue dash inverted triangle—component 1, Red dash—component 2, Green dot dash—component 3, Yellow solid—component 4, Cyan circle dash—component 5, and Pink plus dash—component 6.
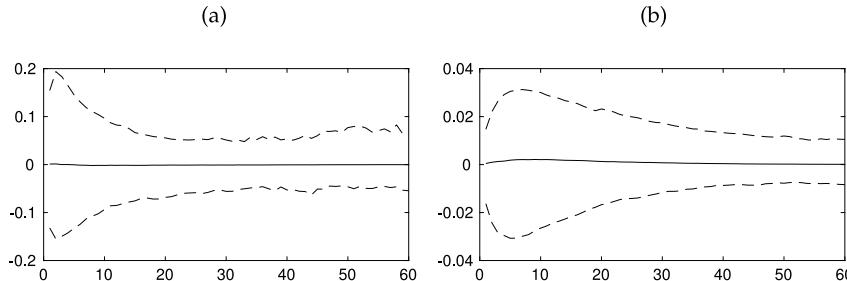
Fig. 7. GIRF of inflation response (a) and GIRF of unemployment rate response (b) to a 1% increase in federal funds rate. We look at 60 months ahead and the shock starts in June 1981. Solid line is the 50th percentile and dashed line the 95% credible interval.
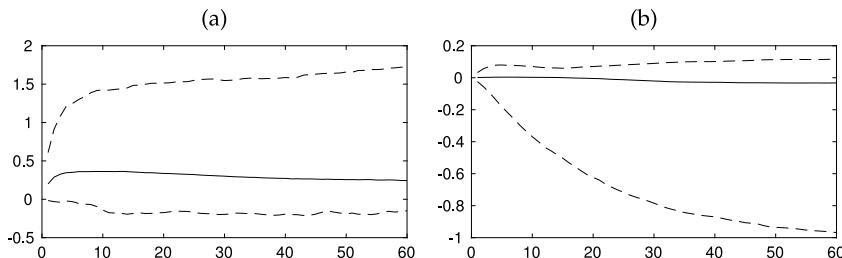
Fig. 8. GIRF of inflation response (a) and GIRF of unemployment rate response (b) to a 1% increase in federal funds rate. We look at 60 months ahead and the shock starts in December 2007. Solid line is the 50th percentile and dashed line the 95% credible interval.

outperforms BVAR(1) at horizon 2, and 4 for all variables and horizon 1 for variable 1.

Tables 5–8 display the log-predictive scores and the RMSEs for the growth rates of the US and UK data respectively. The BayesNP-VAR(1) is on average the model with the better out-of-sample predictive performance for both data sets. The BayesNP-VAR(1) model outperforms the TVP-SV-VAR(1) for all forecasting horizons in both the overall and marginal log-predictive scores, and RMSEs with the US growth rates data. In terms of overall log-predictive scores it is slightly outperformed by the BVAR(3) at horizon 1, but it provides superior overall log-predictive scores for the longer time

horizons of 2 and 4 months than any of the alternative models. The BayesNP-VAR outperforms the BVAR models for most variables at horizons 1, 2 and 4 in log predictive scores and for all variables, except unemployment rate, at all horizons with RMSE. In the case of the UK growth rates, the BayesNP-VAR(1) outperforms the other models for the overall and marginal log-predictive scores and RMSE for all horizons and all variables (with a few exceptions).

Tables 9–12 display the log-predictive scores and RMSEs for the levels of the US and UK data respectively. Once again, the BayesNP-VAR(1) model produces better overall and marginal log-predictive scores and RMSEs for all time horizons when compared to the

Table 3
Log-predictive scores and RMSEs for the simulated VAR(2) data.

Model	Horizon	Joint and marginal scores				RMSEs		
		Overall	Var1	Var2	Var3	Var1	Var2	Var3
BVAR(1)	1	123	49	42	47	0.52	0.58	0.62
	2	135	44	48	50	0.57	0.65	0.69
	4	138	50	52	51	0.70	0.73	0.72
BVAR(2)	1	127	40	42	47	0.55	0.58	0.63
	2	138	44	48	50	0.60	0.66	0.69
	4	141	50	52	51	0.71	0.74	0.74
BVAR(3)	1	129	41	43	50	0.55	0.59	0.67
	2	142	50	50	52	0.63	0.70	0.73
	4	145	51	53	52	0.73	0.78	0.76
BVAR(4)	1	134	43	44	50	0.60	0.62	0.68
	2	150	48	52	53	0.70	0.76	0.74
	4	150	52	54	52	0.78	0.82	0.76
BayesNP-VAR(1)	1	125	39	43	47	0.51	0.57	0.61
	2	133	44	48	50	0.57	0.63	0.69
	4	136	49	51	50	0.69	0.73	0.72
TV-SV-VAR(1)	1	170	47	63	71	0.63	0.68	0.68
	2	177	50	80	78	0.58	0.96	0.82
	4	193	61	87	87	6.17	8.83	6.75

Table 4
Log-predictive scores for simulated threshold VAR(2) data.

Model	Horizon	Joint and marginal scores				RMSEs		
		Overall	Var1	Var2	Var3	Var1	Var2	Var3
BVAR(1)	1	305	74	97	60	1.08	1.83	0.86
	2	335	92	97	67	1.33	1.56	0.93
	4	361	95	109	63	1.43	1.72	0.96
BVAR(2)	1	276	75	82	47	1.42	1.73	0.85
	2	338	94	78	50	1.71	1.46	0.93
	4	341	96	85	51	1.46	1.55	0.97
BVAR(3)	1	291	77	77	63	1.38	1.89	0.97
	2	322	94	74	68	1.57	1.84	1.04
	4	323	94	82	56	1.44	1.69	0.96
BVAR(4)	1	297	78	81	65	1.42	2.08	1.07
	2	313	97	83	68	1.59	2.13	1.19
	4	332	96	82	59	1.46	1.99	1.11
BayesNP-VAR(1)	1	171	67	55	59	1.04	0.75	0.79
	2	204	81	59	70	1.51	0.86	0.83
	4	211	89	61	72	1.91	0.93	0.96
TV-SV-VAR(1)	1	195	69	64	67	1.01	0.75	0.79
	2	229	88	71	79	1.52	0.91	0.97
	4	244	102	76	86	2.39	1.23	1.54

TVP-SV-VAR(1) for both the US and UK data. When compared to the BVAR models, it performs better on the UK data. In the US data, the BayesNP-VAR(1) model outperforms BVAR(1) overall for horizons 1 and 2 but not horizon 4. The BVAR(1) and BayesNP-VAR(1) model each produce better predictions for some variables but there is no clearly superior model for all variables for log predictive scores but BayesNP-VAR(1) outperforms BVAR(1) for all variables with RMSE.

To summarise the BayesNP-VAR(1) outperforms the TVP-SV-VAR(1) for all predictive horizons, in both the simulated and empirical data sets. The non-stationary TVP-SV-VAR accounts for nonlinearity in the conditional mean, and heteroskedasticity in the conditional variance and therefore one would expect it to be better suited to capturing the nonlinear dynamic relationships between variables, leading to better predictive performance. However, for all data sets considered, it is often outperformed by BVAR models. This may be due to fact that the TVP-SV-VAR does not separate the non-linearity in the conditional mean and heteroskedasticity from non-stationarity, which can be a problem when y_t is a stationary process. The transition density of the TVP-SV-VAR depends on the mean μ_t and the covariance Σ_t which are modelled by non-stationary processes. Therefore, when this transition density is not appropriately chosen then the TVP-SV-VAR will not do a good job in approximating the real transition density.

5. Discussion

This paper introduces a new approach to modelling multivariate time series. Using Bayesian nonparametric methods, we have shown how we can express both marginal and transition densities as infinite mixtures, leading to a flexible stationary model that allows for non-linearity in the conditional mean, heteroskedasticity in the conditional variance, and non-Gaussianity. Our empirical results, for both US and UK data, as well as simulated data, indicate that the BayesNP-VAR model outperforms TVP-SV-VAR at all time horizons and shows substantial improvements over BVAR at most time horizons and data sets. The empirical results illustrate that it is useful to allow changes in mixture components to depend on the observed lagged values. However, our model assumes stationarity,

Table 5
Log-predictive scores for growth rates of US data.

Model	Horizon	Joint and marginal scores						
		Overall	FEDFUNDS	INDPRO	LTIR	PAYEMS	PCEPI	UNRATE
BVAR(1)	1	314	13	159	−10	87	96	−23
	2	391	30	157	10	90	98	−5
	4	491	50	152	34	88	94	33
BVAR(2)	1	306	9	160	−10	82	95	−25
	2	392	35	158	14	82	99	−11
	4	540	75	152	50	85	95	15
BVAR(3)	1	301	8	159	−11	81	95	−25
	2	387	36	158	15	81	99	−11
	4	511	72	152	44	83	95	13
BVAR(4)	1	301	8	159	−10	81	95	−25
	2	392	36	158	16	81	98	−11
	4	511	72	152	45	83	95	13
BNP-VAR(1)	1	322	52	109	39	48	52	22
	2	335	51	107	38	47	51	21
	4	321	58	102	36	45	49	20
TV-SV-VAR(1)	1	332	48	120	38	50	63	16
	2	351	54	119	45	51	64	21
	4	367	62	121	50	55	65	26

Table 6
RMSEs for growth rates of US data.

Model	Horizon	RMSEs					
		FEDFUNDS	INDPRO	LTIR	PAYEMS	PCEPI	UNRATE
BVAR(1)	1	0.117	5.91	0.153	0.93	1.71	0.148
	2	0.189	5.92	0.238	0.83	1.88	0.199
	4	0.307	6.04	0.326	0.87	1.77	0.251
BVAR(2)	1	0.130	6.33	0.161	0.73	1.72	0.145
	2	0.233	6.46	0.254	0.66	1.96	0.184
	4	0.390	6.58	0.359	0.70	1.86	0.214
BVAR(3)	1	0.125	6.23	0.163	0.75	1.69	0.144
	2	0.209	6.41	0.249	0.70	1.92	0.182
	4	0.343	6.66	0.323	0.69	1.84	0.204
BVAR(4)	1	0.127	6.17	0.166	0.76	1.69	0.144
	2	0.217	6.39	0.255	0.74	1.92	0.185
	4	0.385	6.65	0.336	0.72	1.91	0.209
BayesNP-VAR(1)	1	0.026	4.94	0.153	0.69	1.67	0.159
	2	0.043	5.05	0.232	0.67	1.81	0.228
	4	0.061	5.10	0.328	0.74	1.76	0.332
TV-SV-VAR(1)	1	0.098	6.39	0.180	0.70	1.99	0.179
	2	0.163	6.24	0.322	0.71	2.41	0.278
	4	0.731	13.99	0.797	2.22	7.00	2.565

which is a strong assumption about macroeconomic data. There may be benefits in terms of predictive power from relaxing this assumption by allowing both marginal and transition densities to vary over time. We will investigate these types of models in future work.

Appendix A. Gibbs sampler

We assume that data $y_1, \ldots, y_T$ are observed and we fit the model with L lags using the following Gibbs sampler. We use the renormalised stick-breaking construction of Griffin (2016) with K atoms as our truncation. This implies a transition density of the form

$$p(y_t | y_{(t-L):(t-1)}) = \frac{\sum_{j=1}^{\infty} w_j k \left(y_{(t-L):t} | \theta_j \right)}{\sum_{j=1}^{\infty} w_j k \left(y_{(t-L):(t-1)} | \theta_j \right)}$$

where $w_j = V_j \prod_{m<j} (1 - V_m)$ and $V_j \overset{iid}{\sim} Be(1, M)$ for $1 \leq k \leq K$ and $\theta_j \overset{iid}{\sim} H$.

Updating μ

The full conditional density of $\mu_{j,i}$ is proportional to

$$\exp \left\{ -\frac{(\mu_{j,i} - \mu_{0,i})^2}{2\sigma_{0,i}^2} \right\} \prod_{t=L+1}^{T} p_K \left(y_t | y_{(t-L):(t-1)} \right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating S

The full conditional density of $S_{j,i}$ is proportional to

$$s_{j,i}^{-(1+\alpha_S)} \exp \left\{ -\zeta_i(\alpha_S - 1)/s_{j,i} \right\} \prod_{t=L+1}^{T} p_K \left(y_t | y_{(t-L):(t-1)} \right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler on the log scale where a normal

Table 7
Log-predictive scores for growth rates of UK data.

Model	Horizon	Joint and marginal scores					
		Overall	CPI	INDPRO	LTIR	STIR	UNRATE
BVAR(1)	1	254	121	183	−1	7	−48
	2	338	120	181	20	23	−25
	4	485	115	173	55	41	23
BVAR(2)	1	235	116	182	−5	4	−55
	2	356	114	180	23	29	−20
	4	647	111	174	72	65	72
BVAR(3)	1	231	115	182	−5	4	−57
	2	346	112	180	24	30	−32
	4	608	111	173	66	64	60
BVAR(4)	1	224	114	182	−4	3	−61
	2	327	111	180	24	29	−40
	4	572	109	173	67	60	27
BayesNP-VAR(1)	1	168	104	145	−12	−42	−67
	2	220	95	144	12	−22	−42
	4	345	93	140	3	−27	−12
TV-SV-VAR(1)	1	353	81	142	49	55	29
	2	368	79	142	54	62	35
	4	389	81	143	61	68	42

Table 8
RMSEs for growth rates of UK data.

Model	Horizon	RMSEs				
		CPI	INDPRO	LTIR	STIR	UNRATE
BVAR(1)	1	3.24	10.19	0.20	0.058	0.076
	2	3.58	10.52	0.35	0.108	0.130
	4	3.51	10.32	0.58	0.205	0.237
BVAR(2)	1	2.87	9.94	0.18	0.123	0.069
	2	2.89	10.22	0.32	0.224	0.115
	4	3.06	10.46	0.56	0.364	0.199
BVAR(3)	1	2.75	9.94	0.18	0.126	0.070
	2	2.75	10.22	0.33	0.248	0.104
	4	3.00	10.42	0.56	0.417	0.169
BVAR(4)	1	2.64	9.96	0.20	0.131	0.062
	2	2.54	10.21	0.37	0.253	0.092
	4	2.70	10.50	0.65	0.509	0.159
BayesNP-VAR(1)	1	2.29	9.93	0.18	0.064	0.089
	2	2.12	9.94	0.30	0.112	0.158
	4	1.99	10.04	0.49	0.198	0.295
TV-SV-VAR(1)	1	2.57	14.33	0.19	0.084	0.087
	2	2.07	15.23	0.32	0.148	0.126
	4	2.11	13.59	0.59	0.542	0.380

proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating ξ

The full conditional density of $\xi_{j,i}$ is proportional to

$$\xi_{j,i}^{\nu/2-1} \exp\left\{-\frac{\nu}{2}\xi_{j,i}\right\} \prod_{t=L+1}^{T} p_K\left(y_t \mid y_{(t-L):(t-1)}\right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler on the log scale where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating ρ

The full conditional density of $\rho_{j,i}$ is proportional to

$$\prod_{t=L+1}^{T} p_K\left(y_t \mid y_{(t-L):(t-1)}\right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler on a logit scale where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating $\rho^\star$

The full conditional density of $\rho_{j,i}^\star$ is proportional to

$$\prod_{t=L+1}^{T} p_K\left(y_t \mid y_{(t-L):(t-1)}\right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler on a logit scale where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating Λ

The full conditional density of $\Lambda_{j,i,k}$ is proportional to

$$\exp\left\{-\frac{1}{2}\Lambda_{j,i,k}^2 \tau_{j,k}\phi_{j,i,k}\right\} \prod_{t=L+1}^{T} p_K\left(y_t \mid y_{(t-L):(t-1)}\right).$$

Table 9
Log-predictive scores for levels of US data.

Model	Horizon	Joint and marginal scores						
		Overall	FEDFUNDS	INDPRO	LTIR	PAYEMS	PCEPI	UNRATE
BVAR(1)	1	327	14	159	−9	98	98	−20
	2	489	30	173	8	118	119	−2
	4	793	46	190	28	160	147	20
BVAR(2)	1	309	9	157	−10	88	94	−25
	2	533	35	177	14	116	130	−7
	4	1057	75	216	47	186	199	26
BVAR(3)	1	301	8	158	−11	82	94	−26
	2	504	36	178	14	104	130	−12
	4	967	69	216	42	170	201	16
BVAR(4)	1	298	8	157	−10	82	94	−27
	2	505	36	177	16	102	132	−13
	4	944	71	213	44	161	200	13
BayesNP-VAR(1)	1	290	−9	210	22	216	178	3
	2	486	15	231	39	257	204	19
	4	899	43	264	47	291	246	27
TV-SV-VAR(1)	1	930	41	280	36	262	298	14
	2	963	51	285	42	266	303	19
	4	996	58	292	47	275	312	25

Table 10
RMSEs for levels of US data.

Model	Horizon	RMSEs					
		FEDFUNDS	INDPRO	LTIR	PAYEMS	PCEPI	UNRATE
BVAR(1)	1	0.112	4.91	0.149	1.73	1.78	0.160
	2	0.215	7.12	0.223	3.38	2.84	0.229
	4	0.400	11.01	0.306	6.67	3.76	0.330
BVAR(2)	1	0.126	5.59	0.161	1.23	1.67	0.143
	2	0.222	8.41	0.251	2.39	2.94	0.194
	4	0.336	13.10	0.335	5.00	4.16	0.259
BVAR(3)	1	0.125	5.73	0.159	0.91	1.68	0.138
	2	0.212	8.97	0.239	1.47	2.97	0.177
	4	0.273	14.82	0.305	2.94	4.17	0.203
BVAR(4)	1	0.128	5.52	0.168	0.86	1.68	0.136
	2	0.211	8.53	0.255	1.33	3.01	0.174
	4	0.318	13.90	0.309	2.49	4.29	0.199
BayesNP-VAR(1)	1	0.031	4.91	0.149	1.17	2.03	0.158
	2	0.057	7.14	0.224	2.23	3.49	0.224
	4	0.093	10.28	0.304	4.40	5.82	0.324
TV-SV-VAR(1)	1	0.032	5.53	0.159	1.43	2.26	0.154
	2	0.056	8.96	0.260	5.85	12.35	0.205
	4	0.131	10.35	0.372	7.52	9.70	0.274

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating ϕ

The full conditional distribution for $\phi_{j,i,k}$ is $\text{Ga}\left((\nu + 1)/2, (\nu + \tau_{j,k}\Lambda_{j,i,k})/2\right)$.

Updating δ

Let $\tau_{j,k}^{(h)} = \prod_{t=1, t \neq h}^{k} \delta_t$. The full conditional distribution for $\delta_{j,1}$ is

$$\text{Ga}\left(a_1 + pq/2, 1 + \sum_{k=1}^{q} \tau_{j,k}^{(1)} \sum_{i=1}^{p} \phi_{j,i,k}\lambda_{j,i,k}^2\right).$$

The full conditional distribution for δ_h, $2 \leq k \leq q$ is

$$\text{Ga}\left(a_2 + p(q - h + 1)/2, 1 + \sum_{k=h}^{q} \tau_{j,k}^{(h)} \sum_{i=1}^{p} \phi_{j,i,k}\lambda_{j,i,k}^2\right).$$

Updating ζ

The full conditional distribution of ζ_i is $\text{Ga}\left(K\alpha_S + \kappa_1, (\alpha_S - 1)\sum_{j=1}^{K}s_{j,i} + \kappa_2\right)$.

Updating V

The full conditional density of V_j is proportional to

$$(1 - V_j)^{M-1} \prod_{t=L+1}^{T} p_K\left(y_t \mid y_{(t-L):(t-1)}\right).$$

The parameter can be updated using an adaptive random walk Metropolis–Hastings sampler on the logit scale where a normal proposal is used whose variance is tuned to have an acceptance rate 0.234.

Updating M

The full conditional distribution of M is $\text{Ga}\left(1 + K, 1 - \sum_{i=1}^{K} \log(1 - V_i)\right)$.

Table 11
Log-predictive scores for levels of UK data.

Model	Horizon	Joint and marginal scores					
		Overall	CPI	INDPRO	LTIR	STIR	UNRATE
BVAR(1)	1	313	128	184	−3	10	−8
	2	620	171	197	16	27	116
	4	1498	305	207	41	49	481
BVAR(2)	1	253	121	183	−6	3	−45
	2	503	154	189	20	28	16
	4	1261	250	194	54	61	239
BVAR(3)	1	236	118	181	−6	4	−53
	2	430	145	188	21	29	−23
	4	972	233	191	57	58	90
BVAR(4)	1	226	116	181	−6	4	−60
	2	399	141	188	20	28	−38
	4	853	230	190	54	56	42
BayesNP-VAR(1)	1	193	124	177	−13	−55	−46
	2	329	155	199	19	−12	−19
	4	561	190	210	59	19	34
TV-SV-VAR(1)	1	617	270	222	47	50	28
	2	641	276	224	52	57	33
	4	673	284	229	59	66	40

Table 12
RMSEs for levels of UK data.

Model	Horizon	RMSEs				
		CPI	INDPRO	LTIR	STIR	UNRATE
BVAR(1)	1	4.09	10.01	0.173	0.281	0.213
	2	7.78	13.81	0.288	0.541	0.408
	4	15.42	16.62	0.459	1.005	0.770
BVAR(2)	1	3.36	10.20	0.164	0.114	0.135
	2	6.26	12.69	0.270	0.272	0.296
	4	12.73	15.11	0.411	0.700	0.643
BVAR(3)	1	3.13	9.71	0.166	0.124	0.094
	2	5.70	12.31	0.280	0.259	0.170
	4	11.74	14.87	0.444	0.497	0.381
BVAR(4)	1	2.99	9.69	0.164	0.120	0.072
	2	5.43	12.22	0.271	0.239	0.118
	4	11.41	14.12	0.425	0.453	0.240
BayesNP-VAR(1)	1	2.81	9.30	0.172	0.039	0.085
	2	4.89	11.96	0.285	0.074	0.149
	4	9.00	12.65	0.442	0.137	0.271
TV-SV-VAR(1)	1	3.34	10.34	0.164	0.061	0.072
	2	4.10	13.26	0.274	0.133	0.113
	4	8.34	12.79	0.474	0.280	0.230

Appendix B. Simulated VAR(2) specification

$$y_t = \begin{pmatrix} 0.13 \\ 0.12 \\ 0.29 \end{pmatrix} + \begin{pmatrix} 0.39 & 0.10 & 0.05 \\ 0.35 & 0.34 & 0.47 \\ 0.49 & 0.24 & 0.24 \end{pmatrix} y_{t-1}$$

$$+ \begin{pmatrix} 0.06 & 0.11 & 0.02 \\ -0.19 & -0.18 & -0.01 \\ -0.31 & -0.13 & 0.09 \end{pmatrix} y_{t-2} + e_t$$

with covariance matrix,

$$\Sigma = \begin{pmatrix} 0.28 & 0.03 & 0.07 \\ 0.03 & 0.29 & 0.14 \\ 0.07 & 0.14 & 0.36 \end{pmatrix}.$$

References

Andrieu, C., Moulines, E., 2006. On the ergodicity properties of some adaptive MCMC algorithms. Ann. Appl. Probab. 16, 1462–1505.

Antoniano-Villalobos, I., Walker, S.G., 2016. A nonparametric model for stationary time series. J. Time Series Anal. 37, 126–142.

Atchadé, Y.F., Fort, G., 2010. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. Bernoulli 16, 116–154.

Atchadé, Y.F., Rosenthal, J.S., 2005. On adaptive Markov chain Monte Carlo algorithms. Bernoulli 11, 815–828.

Auerbach, A.J., Gorodnichenko, Y., 2013. Fiscal multipliers in recession and expansion. In: Alesina, A., Giavazzi, F. (Eds.), Fiscal Policy after the Financial Crisis, no. 63–98. University of Chicago Press.

Baum, A., Koester, G.B., 2011. The Impact of Fiscal Policy on Economic Activity Over the Business Cycle: Evidence from a thershold VAR analysis, Discussion paper, Series 1 Economic Studies N03/2011. Deutsche Bundesbank.

Beaudry, P., Koop, G., 1993. Do recessions permantly change output?. J. Monet. Econ. 31, 149–163.

Bhattacharya, A., Dunson, D.B., 2011. Sparse Bayesian infinite factor models. Biometrika 98 (2), 291–306.

Carriero, A., Clarke, T.E., Marcellino, M., 2015. Bayesian VARs: Specification choices and forecasting accuracy. J. Appl. Econometrics 30, 46–73.

Chauvet, M., 1998. An econometric characterization of business cycle dynamics with factor structure and regime switching. Internat. Econom. Rev. 39, 969–996.

Chib, S., 1996. Calculating posterior distributions and model estimates in Markov mixture models. J. Econometrics 75, 79–97.

Cogley, T., Sargent, T.J., 2001. Evolving post world war II US inflation dynamics. In: NBER Macroeconomics Annual, vol. 16. pp. 331–338.

Cogley, T., Sargent, T.J., 2005a. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. Rev. Econ. Dyn. 8, 262–302.

Dahl, C.M., Gonzalez-Rivera, G., 2003a. Identifying nonlinear components by random fields in the US GNP growth. Implications and shape of the business Cycle. Stud. Nonlinear Dyn. Econ. 7.

Dahl, C.M., Gonzalez-Rivera, G., 2003b. Testing of neglected nonlinearity in regression models based on the theory of random fields. J. Econometrics 114, 141–164.

Diebold, F.X., Rudebusch, G.D., 2001. Five questions about business cycles. FRBSF Econ. Rev. 1–15.

Gambacorta, L., Hofmann, B., Peersman, G., 2014. The effectiveness of unconventional monetary policy at the zero lower bound: A cross-country analysis. J. Money Credit Bank. 46, 615–642.

Geweke, J., Keane, M., 2007. Smoothly mixing regressions. J. Econometrics 138, 252–290.

Granger, C., Terasvirta, T., 1994. Modelling nonlinear economic relationships. Int. J. Forecast. 10, 169–171.

Griffin, J.E., 2016. An adaptive truncation method for inference in Bayesian nonparametric models. Stat. Comput. 26, 423–441.

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

Hamilton, J.D., 2001. A parametric approach to flexible nonlinear inference. Econometrica 69, 537–573.

Hansen, B.E., 1992. The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP. J. Appl. Econometrics 7, 61–82.

Härdle, W., Tsybakov, A., Yang, L., 1998. Nonparametric vector autorgression. J. Statist. Plann. Inference 68, 221–245.

Hjort, N.L., Holmes, C., Müller, P., Walker, S.G. (Eds.), 2010. Bayesian Nonparametrics, first ed. In: Statistic and Probabilistic Mathematics, Cambridge University Press.

Hubrich, K., Teräsvirta, T., 2013. Thresholds and smooth transitions in vector autoregressive models. In: Fomby, T.B., Kilian, L., Murphy, A. (Eds.), VAR Models in Macroeconomics-New Developments and Applications: Essays in Honor of Christopher A. Sims. In: Advances in Econometrics, vol. 32, Emerald Group Publishing Limited.

Hyvärinen, A., Zhang, K., Shimizu, S., Hoyer, P.O., 2010. Estimation of a structural vector autoregression model using non-Gaussianity. J. Mach. Learn. Res. 11, 1709–1731.

Jacobs, R., Jordan, M., Nowlan, S., Hinton, G., 1991. Adaptive mixtures of local experts. Neural Comput. 3, 79–87.

Jeliazkov, I., 2013. Nonparametric vector autoregressions: Specification, estimation and inference. In: Fomby, T.B., Kilian, L., Murphy, A. (Eds.), VAR Models in Macroeconomics-New Developments and Applications: Essays in Honor of Christopher A. Sims. In: Advances in Econometrics, vol. 32, Emerald Group Publishing Limited.

Jordan, M., Jacobs, R., 1994. Hierarchical mixtures of experts and the EM algorithm. Neural Comput. 6, 181–214.

Kim, C.J., Morley, J., Piger, J., 2005. Nonlinearity and the permanent effects of recessions. J. Appl. Econometrics 20, 291–309.

Kim, C.J., Nelson, C.R., 1999. Has the us economy become more stable? a Bayesian approach based on the business cycle. Rev. Econ. Stat. 81, 608–616.

Koop, G., Korobilis, D., 2010. Bayesian multivariate time series methods for empirical macroeconomics. Found. Trends Econ. 3, 267–358.

Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impluse response analysis in nonlinear multivariate models. J. Econometrics 74, 119–147.

Lanne, M., Lütkepohl, H., 2010. Structural vector autoregressions with nonnormal residuals. J. Bus. Econom. Statist. 28, 159–168.

Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196, 288–304.

Lucas, R.E., 1980. Methods and problems in business cycle theory. J. Money Credit Bank. 12, 696–715.

Matthes, C., Barnichon, R., 2015. Measuring The Nonlinear Effects of Monetary Policy, Number 49 in 2015 Meeting Papers. Society of Economic Dynamics.

Moneta, A., Entner, D., Hoyer, P.O., Coad, A., 2013. Causal inference by independent component analysis: theory and applications. Oxf. Bull. Econ. Stat. 75, 705–730.

Norets, A., Pati, D., 2017. Adaptive Bayesian estimation of conditional densities. Econometric Theory 33, 980–1012.

Pagan, A., 1997. Towards an understanding of some business cycle characteristics. Aust. Econ. Rev. 30, 1–15.

Pesaran, M.H., Potter, S.M., 1997. A floor and ceiling model of US output. J. Econom. Dynam. Control 21, 661–695.

Potter, S.M., 1995. A non linear approach to US GNP. J. Appl. Econometrics 10, 109–125.

Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy. Rev. Econom. Stud. 72, 821–852.

Ravn, M.O., Sola, M., 2004. Asymmetric effects of monetary policy in the US. Fed. Reserve Bank St. Louis Rev. 86, 41–60.

Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statist. Sinica 4 (2), 639–650.

Sims, C., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

Sims, C., Zha, T., 2006. Were there regime switches in us monetary policy?. Am. Econ. Rev. 96, 54–81.

Sörensen, B.E., Wu, L., Yosha, O., 2001. Output fluctuations and fiscal policy. Eur. Econ. Rev. 45, 1271–1310.

Stock, J., Watson, M., 1996. Evidence of structural instability in macroeconomic time series relations. J. Bus. Econom. Statist. 14, 11–30.

Stock, J., Watson, M., 1999. Handbook of Macroeconomics, vol. 1, Part A. Elsevier, Chapter 1: Business cycle fluctuations in US macroeconomic time series.

Stock, J., Watson, M., 2001. Vector autoregressions. J. Econ. Perspect. 15, 101–116.

Stock, J., Watson, M., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20, 147–162.

Teräsvirta, T., 1994. Specification, estimation and evaluation of smooth transition autoregressive models. J. Amer. Statist. Assoc. 89, 208–218.

Villani, M., Kohn, R., Giordani, P., 2009. Regression density estimation using smooth adaptive Gaussian mixtures. J. Econometrics 153, 155–173.

Villani, M., Kohn, R., Nott, D., 2012. Generalized smooth finite mixtures. J. Econometrics 171, 121–133.

Weise, C.L., 1999. The asymmetric effects of monetary policy: A non linear vector autoregression approach. J. Money Credit Bank. 31, 85–108.