Bayesian nonparametric vector autoregressive models

Maria Kalli and Jim E. Griffin *

September 11, 2015

Abstract

Vector autoregressive (VAR) models are the main work-horse model 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 to 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 USA and Eurozone 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 predictions, we show that BayesNP-VAR predictively outperforms competing models.

Keywords: Vector Autoregressive Models; Dirichlet Process Prior; Infinite Mixtures; Markov chain Monte Carlo; Inflation. JEL codes: C11, C15, C52, C53, and C58.

*Maria Kalli, Canterbury Christ Church University Business School, Canterbury Christ Church University, Canterbury, UK (email: maria.kalli@canterbury.ac.uk), Jim Griffin, School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK (email: J.E.Griffin-28@kent.ac.uk).
1 Introduction

Introduced by Sims (1980), vector autoregressive (VAR) models provide a systematic way of capturing the complex dynamics of multiple time-series and their interactions. In its basic form, the 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,

$$y_t = B_1 y_{t-1} + \ldots B_L y_{t-L} + \epsilon_t \text{ for } t = 1, \ldots, T$$  \hspace{1cm} (1)

where $\{B_l\}_{l=1}^{L}$ are $(p \times p)$-dimensional matrices of unknown coefficients, and $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{p,t})'$ is a $(p \times 1)$-dimensional innovation vector with distribution $N(0, \Sigma)$.

VAR models are widely used in the area of macroeconomics and have become the standard benchmark model for the analysis of dynamic economic problems. This is due to the simple linear representation of the variables’ joint dynamic behaviour and the straightforward, easily interpreted statistical properties of the model in (1). One of the main drivers behind their popularity is the need to capture key features of the business-cycle\(^1\). The majority of empirical macroeconomic studies are designed to adequately account for the recurrent and persistent nature of the business cycle as well as the difference in the length of expansion and contraction periods\(^2\). For a detailed review, see Lucas (1977, 1980), Pagan (1997), Stock and Watson (1999), and Diebold and Rudebusch (2001). VAR models, through the use of impulse response functions, and forecast error variance decompositions, assist in analysing the effects of monetary and fiscal policy on the business-cycle and therefore lead to a better understanding of its characteristics. VAR models also provide flexible (point or density) forecasts as they can be made conditional on potential future paths of the other macroeconomic variables. This can be useful in forecasting overall economic activity, and future turning points (peaks and troughs)\(^3\) of the business cycle.

\(^1\)The short-run fluctuations in aggregate economic activity around its long-run growth path.

\(^2\)Expansions are fairly long whereas contractions are fairly short.

\(^3\)The maximum and minimum points of aggregate economic activity.
There have been several main criticisms of VAR models. Their rich parameterisation is attached to the risk of overfitting the data leading to imprecise inference, while the interdependence of the innovations ($\epsilon_t$’s) leads to identification issues. Often, the linearity, stationarity, Gaussian innovations and constant conditional mean and variance of the model in (1) can be considered unrealistic. For example, empirical evidence suggests that the relationship between macroeconomic variables may not be linear and the nature of the shocks may not be Gaussian, see Granger and Terasvirta (1994), and Weise (1999). Some studies on the effects of monetary policy shocks on the business cycle conclude that contractionary and expansionary shocks have different effects on output. The former having a large and significant negative impact on output and the latter having very small and insignificant impact, see Cover (1992) and Weise (1999). The more recent work of Ravn and Sola (2004) and Barnichon and Matthes (2014) also finds that the size of the expected shock has different effects on output, as well as the shocks’ timing (i.e. the phase of the business cycle). In terms of fiscal policy, Giavazzi et al. (2000), Sörensen et al. (2001), Auerbach and Gorodnichenko (2011), Baum and Koester (2011), and Gambacorta et al. (2014) find that the effect of a shock is non-linear, and depends on the phase of the business cycle, and the timing and size of the shock.

Luckily, the model in (1) is very versatile, and in the past couple of decades these criticisms have been addressed using adaptations which account for time varying parameters, stochastic volatility, structural breaks, regime switching and threshold crossing behaviour to simply name a few. Hamilton (1989) introduced Markov-switching regression to characterise the changes in the parameters of the autoregressive process, thus modelling changes in regime. Since then the literature on models to account for regime shifts in the mean, variance or dynamics has 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. These
models describe scenarios in which the dynamic behaviour of the variables can be modelled by a small number of linear states. An observable transition variable determines the regime that generates the next observation. VTAR uses an indicator function to model the change in states, whereas VSTAR models this change using a continuous function which allows the dynamic behaviour of the variables to change smoothly between states. For a comprehensive survey see Hubrich and Teräsvirta (2013). An alternative class of extensions to model (1) are time-varying parameter models with or without stochastic volatility. In these models, a VAR model is assumed whose conditional mean or variance is allowed to vary over time. Notable work on these models can be found 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 all the aforementioned models.

In contrast to parametric VARs with Gaussian innovations or time-varying parameter versions, the literature on the use of nonparametric methods and non-Gaussian innovations in VAR models is not as rich, and this is where our contribution lies. Härdle et al. (1998) proposed the vector conditional heteroscedastic autoregressive nonlinear (CHARN) 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. He used his model to address nonlinearity in the inflation-unemployment trade-off. Dahl and Gonzalez-Rivera (2003a,b) extended his model to non-Gaussian random fields and used their methods to analyse the evolution of industrial production for a sample of OECD countries, and the evolution of GNP growth in the US. Finally, 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 a past realisations. The use of non-Gaussian innovations ($\epsilon_i$’s) is fairly recent and linked to structural VAR models where it can be used to address the identification issues attached to Gaussian innovations. Hyvärinen et al. (2010),
Moneta et al. (2013), and Lanne et al. (2015) use independent component analysis (ICA). They assume the $\epsilon_t$'s to be independent and non-Gaussian and represent the residuals (obtained when estimating the VAR model) as linear mixtures of them, connected by some 'mixing matrix', $\Gamma$. Lanne and Lütkepohl (2010) model the innovations using a mixture of two Gaussian distributions, whereas Jeliazkov (2013) uses the student t-distribution.

In this paper, we develop a VAR model that allows for nonlinearity in the conditional mean, heteroscedasticity in the conditional variance, and non-Gaussian innovations. We follow a Bayesian approach, that differs to that of previous studies. We do not assign priors to the VAR parameters, instead we directly model the stationary and transition densities using a Bayesian nonparametric approach. Bayesian nonparametric models place a prior on an infinite dimensional parameter space and adapt their complexity to the data. A more appropriate term is infinite capacity models, emphasising the crucial property that they allow their complexity (i.e., the number of parameters) to grow as more data is observed; in contrast, finite-capacity models assume a fixed complexity. Hjort et al. (2010) is a recent book length review of Bayesian nonparametric methods. Our approach defines a prior on the stationary and transition densities that has full support, that is any stationary and transition density can be arbitrarily well represented by the prior.

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 of 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 provides an empirical illustration where the model is applied to USA and Eurozone macroeconomic time series. In this section we also compare the BayesNP-VAR to the parametric BVAR and TVP-VAR. Section 5 summarises our findings and conclusions.
2 The Bayesian non-parametric vector autoregressive (BayesNP-VAR) model

Our Bayesian nonparametric model is constructed by expressing the joint distribution of \( y_t \) and its \( L \) lags \( y_{t-1}, \ldots, y_{t-L} \) as an infinite mixture. Bayesian infinite mixture models were popularised by Lo (1984). In the Dirichlet process mixture (DPM) model, an unknown density \( p(\cdot) \) is modelled as

\[
p_{v,\theta}(y) = \sum_{j=1}^{\infty} w_j k(y; \theta_j)
\]

where \( k(y; \theta) \) is a continuous density function for \( y \) with parameters \( \theta \), the cluster locations are \( \theta_j \sim H \). The weights, which are independent of the cluster locations, are defined using the stick-breaking representation of the Dirichlet process, see Sethuraman (1994), with \( w_1 = v_1 \), \( w_j = v_j \prod_{m<j}(1 - v_m) \), and \( v_j \sim \text{Be}(1, M) \). For more on the long history of the stick-breaking notion of constructing infinite dimensional priors see Halmos (1944), Freedman (1963), Kingman (1974), and Ishwaran and James (2001). We will write \( \theta = (\theta_1, \theta_2, \ldots) \) to group the cluster locations. The choice of \( H \) determines the likely location of the clusters and \( M \) controls the relative size of the weights. The expectation of the \( j^{th} \) weight is \( \mathbb{E}[w_j] = \frac{M^{j-1}}{(M+1)^j} \) and so, as \( M \) increases, the average size of the weights becomes smaller and the number of components with non-negligible weights becomes larger. We discuss how to choose \( M \) later in this section.

Antoniano-Villalobos and Walker (2014) describe a Bayesian nonparametric approach to inference for a univariate stationary time series \( y_1, \ldots, y_T \). They showed that their approach defines a prior which 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 where \( \{y_{t,i}\}_{i=1}^p \) is a \( p \)-dimensional vector of macroeconomic variables measured at time \( t \). We refer to our model as the Bayesian nonparametric VAR (BayesNP-VAR) model.
We begin by assuming that the transition density depends on \(L\) lags and define a DPM model for the joint distribution \(y_{(t-L):t}\) where \(y_{s:t} = (y_s, \ldots, y_t)\) for \(s \leq t\). The model assumes that

\[
p(y_{(t-L):t}) = \sum_{j=1}^{\infty} w_j \, k(y_{(t-L):t} | \theta_j)
\]

where the cluster locations \(\theta_j \sim H\), the cluster weights \(w_j\) are defined as in (2), and \(k(y_{(t-L):t} | \theta_j)\) is a \(((L + 1)p)\)-dimensional probability density function which does not depend on \(t\). This leads to the \(p\)-dimensional stationary density of \(y_t\),

\[
p(y_t) = \sum_{j=1}^{\infty} w_j \, k(y_t | \theta_j),
\]

where \(k(y_t | \theta_j)\) is the marginal density of \(k(y_{(t-L):t} | \theta_j)\). The transition density will also be an infinite mixture and has the form

\[
p(y_t | y_{(t-L):(t-1)}) = \frac{p(y_{(t-L):t})}{p(y_{(t-L):(t-1)})} = \frac{\sum_{j=1}^{\infty} w_j \, k(y_{(t-L):t} | \theta_j)}{\sum_{j=1}^{\infty} w_j \, k(y_{(t-L):(t-1)} | \theta_j)} = \sum_{j=1}^{\infty} \omega_j(y_{(t-L):(t-1)}) \, k(y_t | y_{(t-L):(t-1)}, \theta_j)
\]

where \(\omega_j(y_{(t-L):(t-1)}) = \frac{w_j \, k(y_{(t-L):(t-1)} | \theta_j)}{\sum_{k=1}^{\infty} w_k \, k(y_{(t-L):(t-1)} | \theta_k)}\) is the weight of the \(j^{th}\) component which depends on previous lags, and \(k(y_t | y_{(t-L):(t-1)}, \theta_j)\) is the transition density of the \(j^{th}\) component. Equations (4) and (5) provide a flexible mixture model specification of both the stationary and transition density. The transition density is an infinite mixture model whose components are the transition density of the components of (3). The weight for the \(j^{th}\) component varies with the lagged values of \(y_t\) and is proportional to the value of the stationary density associated with the components of (3). This allows the infinite mixture model to favour different components as the lagged values change. For example, this would allow different component transition densities to be favoured in expansionary or contractionary periods with those periods indicated by the values of the lagged variables.

To complete the BayesNP-VAR model we need to choose a stationary time series model for \(k(y_t | y_{(t-L):(t-1)}, \theta_j)\), the transition density within the \(j^{th}\) component for which \(k(y_{(t-L):t} | \theta_j)\) and \(k(y_{(t-L):(t-1)} | \theta_j)\), the joint densities of that component, can be easily derived. We choose
a dynamic factor model with \( q \) factors and parameters \( \theta_j = (\mu_j, S_j, \Lambda_j, \rho_j^*, \mu_j, \xi_j) \) within each component which leads to a sparse decomposition of the covariance matrix. For ease of exposition we will drop \( j \) from \( \theta_j \). The form of the dynamic factor model with \( q \) factor is

\[
S^{-1}(y_t - \mu) = \Lambda x_t + \epsilon_t,
\]

\[
x_t = \Gamma x_{t-1} + \nu_t, \quad \nu_t \sim N(0, \Delta)
\]

\[
\epsilon_t = \Phi \epsilon_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Psi)
\]

where \( \mu \) is the \((p \times 1)\)-dimensional stationary mean of \( y_t \), \( S = \text{diag}(s_1, \ldots, s_p) \) is a \((p \times p)\)-dimensional scaling matrix, \( \Lambda \) is a \((p \times q)\)-dimensional vector of factor loadings, \( x_t = (x_{t,1}, \ldots, x_{t,q}) \) is \((q \times 1)\)-dimensional vector of factor scores at time \( t \), and \( \epsilon_t \) is a \((p \times 1)\)-dimensional vector of independent errors at time \( t \). For the AR(1) representation of the factor scores \( x_t \) we assume that \( \Gamma = \text{diag}(\rho_1, \ldots, \rho_q) \) and \( \Delta = \text{diag}(1 - \rho_1^2, \ldots, 1 - \rho_q^2) \) which are both \((q \times q)\)-dimensional matrices. This implies that the stationary distribution of \( x_{z,t} \) for \( z = 1, \ldots, q \) is a standard normal distribution. For the AR(1) representation of the innovations, \( \epsilon_t \) we assume that \( \Phi = \text{diag}(\rho_1^*, \ldots, \rho_p^*) \) and \( \Psi = \text{diag}(\xi_1^{-1}(1 - \rho_1^*2), \ldots, \xi_p^{-1}(1 - \rho_p^*2)) \), which are both \((p \times p)\)-dimensional matrices, leading to \( \epsilon_{i,t} \) having a zero-mean normal stationary distribution with variance \( \xi_i^{-1} \) for \( i = 1, \ldots, p \). Under this dynamic factor specification the stationary distribution of \( y_t \) is

\[
y_t \sim N(\mu, S(\Lambda \Lambda^T + A)S^T)
\]

where \( A = \text{diag}(\xi_1^{-1}, \ldots, \xi_p^{-1}) \) is a \((p \times p)\)-dimensional matrix and together with \( \Lambda \Lambda^T \) divide the correlation matrix into dependent and idiosyncratic parts.

The joint distribution \( y_{(t-L):((t-L)+K)} \sim N(1_{K+1} \otimes \mu, Q + B) \) for \( K = 1, \ldots, L \) where \( \otimes \) is the Kronecker product and \( 1_k \) is a \((k \times 1)\)-dimensional vector of 1’s. The Kronecker product
$1_{K+1} \otimes \mu$ results in a $((K + 1)p) \times 1$ stacked vector,

$$
\begin{pmatrix}
\mu \\
\mu \\
\vdots \\
\mu
\end{pmatrix}
$$

where $\mu$ is a $p$-dimensional mean vector. $Q$ is a $((K + 1)p) \times ((K + 1)p)$-dimensional matrix with entries

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

where $\Sigma_l$ is the variance matrix for the $l^{th}$ lag, with entries,

$$
\Sigma_l = 
\begin{pmatrix}
\frac{S_1 \rho_{l}^{1}}{\xi_1} & 0 & \ldots & 0 \\
0 & \frac{S_2 \rho_{l}^{2}}{\xi_2} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \frac{S_p \rho_{l}^{p}}{\xi_p}
\end{pmatrix},
$$

for $l = 1, \ldots, K$ and

$$
B = \sum_{h=1}^{q}(1_{L-1} \otimes (SA_z))(P_z \otimes 1_p)(1_{L-1} \otimes (SA_z))^T,
$$

where $P_z$ is a $((K + 1) \times (K + 1))$-dimensional matrix of correlations between the $i^{th}$ and $k^{th}$ variable for the $z^{th}$ factor at different lags, i.e.

$$
P_z = 
\begin{pmatrix}
1 & \rho_z & \ldots & \rho_z^{K} \\
\rho_z & 1 & \ldots & \rho_z^{K-1} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_z^{K} & \rho_z^{K-1} & \ldots & 1
\end{pmatrix}.
Using (5) we can now define \( k(y_t|y_{(t-L):(t-1)}, \theta) \), the transition density, which has an explicit form for the stationary distribution. Then for the \( j^{th} \) component, of the BayesNP-VAR with \( L \) lags the transition is

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

where the location \( \theta_j \overset{iid}{\sim} H \) and the weight \( w_j = v_j \prod_{m<j} (1 - v_m) \), and \( v_j \overset{iid}{\sim} \text{Be}(1, M) \).

The last step to complete the BayesNP-VAR model is to choose \( H \), the prior of the locations \( \theta_j = (\mu_j, S_j, \Lambda_j, \rho_j^*, \rho_j, \xi_j) \). We assume that the parameters are \textit{a priori} independent so that the density of \( H \) is \( h(\mu_j, S_j, \Lambda_j, \rho_j^*, \rho_j, \xi_j) = h_\mu(\mu_j) \times h_S(S_j) \times h_\Lambda(\Lambda_j) \times h_{\rho^*}(\rho_j^*) \times h_\rho(\rho_j) \times h_\xi(\xi_j) \). For the prior density of the stationary mean vector of \( y_t \), we choose \( h_\mu(\mu) = \text{N}(\mu|\mu_0, \Sigma_0) \). Both parameter can be chosen with prior information but we use the data dependent choices \( \mu_0 = \bar{y}_t \) and \( \Sigma_0 = 1.5^2 \bar{\Sigma} \), where \( \bar{\Sigma} \) is the covariance matrix of \( y_t \) in this paper. We choose \( h_\mu(\rho) = \prod_{i=1}^{\nu} U(\rho_i|0,1) \) where \( U(\rho|0,1) \) represents the density of a uniform distribution between 0 and 1. We choose \( h_{\rho^*}(\rho^*) = \prod_{i=1}^{\nu} U(\rho_i^*|0,1) \). The choice of independent uniform priors is driven by the fact that the AR parameters should be positive between 0 and 1. For the \((p \times p)\)-dimensional scaling matrix, \( S = \text{diag}(s_1, s_2, \ldots, s_p) \), we choose a hierarchical prior. Each diagonal element of \( S \) is conditionally independent with \( s_i^{-1} \sim \text{Ga}(a_s, \mu_{s,i}(a_s - 1)) \) (so that \( E(s_i) = \mu_{s,i} \)) and \( \mu_{s,i} \overset{iid}{\sim} \text{Ga}(1, 5) \). This allows each component to have different scales for each variable but implies similar scales for each variable across the components. To avoid specifying the number of factors \textit{a priori}, we use the multiplicative gamma process shrinkage prior of Bhattacharya and Dunson (2011). Under this prior, \( h_\Lambda(\Lambda) = \prod_{i=1}^{p} \prod_{z=1}^{q} N(\Lambda_{i,z}|0, \phi_{i,z}^{-1} \tau_z) \) where \( \phi_{i,z} \sim \text{Ga}(\nu/2, \nu/2) \) and \( \tau_z = \prod_{i=1}^{z-1} \delta_z \) with \( \delta_1 \sim \text{Ga}(1,1) \) and \( \delta_z \sim \text{Ga}(3,1) \) for \( z \geq 2 \). The \( \delta_z \)'s are independent, and the \( \tau_z \)'s are viewed as the global shrinkage parameters of the columns, while \( \phi_{i,z} \)'s are the local shrinkage parameters for the \( z^{th} \) column. As value of \( \delta_z \) increases, so does the value of \( \tau_z \) favour smaller values of \( \Lambda_{i,z} \). Finally, the prior density of the variances of the innovations is \( h_\xi(\xi) = \prod_{i=1}^{p} \text{Ga}(\xi_i|\nu/2, \nu/2) \).
3 Inference in Bayesian NP-VAR

3.1 Posterior computation

We make inference using the transition in (7) to define the likelihood function as

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

Inference is complicated by the infinite sum in both the numerator and denominator (unlike the standard infinite mixture model). Antoniano-Villalobos and Walker (2014) 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 (2014) which adaptively truncates the infinite sum in the numerator and denominator and tends to avoid large truncation errors in the posterior. In the method, an infinite sequence of posteriors for truncated versions of the nonparametric model with a decreasing level of truncation is created. A sequential Monte Carlo method is used to efficiently simulate the sequence of posteriors. Simulation from the sequence is stopped if the differences in successive posteriors, calculated using the effective sample size, is small. We use a truncation of the infinite model in (3) leading to a truncated transition density which has the form

$$p_K(y_t|y_{(t-L):(t-1)}) = \frac{\sum_{j=1}^{K} w_j k(y_{(t-L):(t-1)}|\theta_j)}{\sum_{j=1}^{K} w_j k(y_{(t-L):(t-1)}|\theta_j)}$$

where \(w_j = V_j \prod_{m<j}(1 - V_m)\) and \(V_k \sim \text{Be}(1, M)\) and \(\theta_j \sim H\) which can be used to define a sequence of posterior of the form

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

where \(\eta_{1:K} = (\phi_{1:K}, \delta, V_{1:K}, \mu_S, M)\). An MCMC algorithm is run for \(\pi_{K_0}(\theta_{1:K_0}, \eta_{1:K_0}|y)\) where \(K_0\) is a user-defined starting value. 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 \( \sigma_t^2 \) 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 \( \alpha_t \) is the acceptance probability in the Metropolis-Hastings algorithm at the \( t \)-th 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. The proposed value \( \lambda' \) of a parameter \( \lambda \) at the \( t \)-th iteration is

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

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

Once we have simulated a sample using the MCMC sampler from \( \pi_{K_0}(\theta_{1:K_0}, \eta_{1:K_0}|y) \) which will be denoted \( (\theta^{(1)}_{1:K_0}, \eta^{(1)}_{1:K_0}), \ldots, (\theta^{(N)}_{1:K_0}, \eta^{(N)}_{1:K_0}) \) and set \( K = K_0 \). The following algorithm is run with \( K \)-th step,

1. Set \( K = K + 1 \).
2. Simulate \( \left(\theta_{K+1}^{(i)}, \phi_{K+1}^{(i)}, V_{K+1}^{(i)}\right) \) from their prior distribution for \( i = 1, \ldots, N \).
3. Evaluate
   \[
   \delta_i = \frac{\pi_{K+1}(\theta^{(i)}_{1:(K+1)}, \eta^{(i)}_{1:(K+1)}|y)}{\pi_K(\theta^{(i)}_{1:K}, \eta^{(i)}_{1:K}|y)}, \quad i = 1, \ldots, N
   \]
4. Evaluate
   \[
   \text{ESS}_{K+1} = \frac{(\sum_{i=1}^{N} \delta_i)^2}{\sum_{i=1}^{N} \delta_i^2}
   \]
5. If \( \text{ESS}_i < cN \) then generate \( N \) values where \( \left(\theta_{K+1}^{(i)}, \phi_{K+1}^{(i)}\right) \) is generated with probability proportional to \( \delta_i \). Set \( \delta_i = 1 \) for \( i = 1, \ldots, N \) and run one iteration of the MCMC sampler updating \( \left(\theta^{(i)}_{1:(K+1)}, \eta^{(i)}_{1:(K+1)}\right) \) from \( \pi_{K+1}(\theta_{1:(K+1)}, \eta_{1:(K+1)}|y) \) for \( i = 1, \ldots, N \).
6. Let $D_{K+1} = |ESS_{K+1} - ESS_K|$. If $D_{K+1} \leq \epsilon$, $D_K \leq \epsilon$ and $D_{K-1} \leq \epsilon$ terminate. Otherwise return to step 1.

3.2 Posterior summaries

Our Bayesian nonparametric VAR model is non-normal and non-linear and so many methods for reporting inferences in VAR model cannot be directly extended. We are able to report the joint stationary distribution of a subset, $y_A$, of variables

$$p(y_{t,A}) = \frac{1}{N} \sum_{k=1}^{N} \left[ \sum_{j=1}^{K} w_j^{(k)} k \left( y_{t,A} \mid \theta_j^{(k)}, \eta_j^{(k)} \right) \right]$$

where $\left( \theta_1^{(1)}, \eta_1^{(1)} \right), \ldots, \left( \theta_N^{(1)}, \eta_N^{(1)} \right)$ is a sample from the posterior of the model truncated to $K$ components (as chosen using the adaptive truncation algorithm). The conditional distributions of one set of variables given a disjoint set of variables, $y_B$, is

$$p(y_{t,A} \mid y_{t,B}) = \frac{1}{N} \sum_{k=1}^{N} \left[ \sum_{j=1}^{K} w_j^{(k)} k \left( y_{t,A} \mid y_{t,B}, \theta_j^{(k)}, \eta_j^{(k)} \right) \right]$$

We can also calculate the joint distribution of a subset of variables $y_A$ conditional on a subset of lagged variables $y_C$ (which may share variables with $y_A$) using the expression

$$p(y_{t,A} \mid y_{t-1,C}) = \frac{1}{N} \sum_{k=1}^{N} \left[ \sum_{j=1}^{K} w_j^{(k)} k \left( y_{t,A} \mid y_{t-1,C}, \theta_j^{(k)}, \eta_j^{(k)} \right) k \left( y_{t-1,C} \mid \theta_j^{(k)}, \eta_j^{(k)} \right) \right]$$

For these distributions we can calculate summaries such as the expectation of $y_{t,A}$ given $y_{t-1,C}$ which is

$$p(y_{t,A} \mid y_{t-1,C}) = \frac{1}{N} \sum_{k=1}^{N} \left[ \sum_{j=1}^{K} w_j^{(k)} E \left( y_{t,A} \mid y_{t-1,C}, \theta_j^{(k)}, \eta_j^{(k)} \right) k \left( y_{t-1,C} \mid \theta_j^{(k)}, \eta_j^{(k)} \right) \right]$$

Our Bayesian NP-VAR model is a mixture and so can be expressed in terms of allocation variables $s_1, \ldots, s_T$ as

$$p(y_t \mid s_t) = k(y_t \mid y_{(t-L):(t-1)}, \theta_{s_t}), \quad p(s_t = j) = \omega_j (y_{(t-L):(t-1)}), \quad t = 1, \ldots, T.$$
The allocation variables allow us to cluster observations which have similar relationships between the variables. A sample of $s_1, \ldots, s_T$ using the truncated posterior derived in Section 3.1 can be easily sampled using the full conditional distribution of $s_t$

$$p(s_t|y) \propto \omega_j \left( y_{(t-L):(t-1)} \right) k \left( y_t| y_{(t-L):(t-1)}, \theta_{s_t} \right) \propto w_j k \left( y_{(t-L):t}, \theta_{s_t} \right), \quad t = 1, \ldots, T.$$  

The impulse response function is the standard tool for underlying the persistence and effect of shocks in macroeconomic time series. We use the Generalised Impulse Response Function (GIRF) of Koop et al. (1996) which is defined as

$$\text{GI}_\gamma (n, \nu_t, F_{t-1}) = E [Y_{t+n}|\nu_t, F_{t-1}] - E [Y_t|F_{t-1}]$$

where $\nu_t$ represents an arbitrary shock at time $t$ and $F_{t-1}$ represents the history of $Y_s$ to time $t - 1$. Koop et al. (1996) describe several variations of their measures including

$$\text{GI}_\gamma (n, \nu_t, F_{t-1}) = E [Y_{t+n}|\nu_t] - E [Y_{t+n}]$$

where there is no conditioning on the past history which can be easily calculated for our model since the stationary distribution is easily available.

4 Empirical Examples and Results

The macroeconomic data sets used to illustrate the Bayesian nonparametric VAR model include seasonally adjusted series, taken on a monthly frequency. We use monthly rather than quarterly data because we want data sets with sufficiently many observations, while avoiding long sample periods with many known structural breaks. We therefore chose March 1995 to December 2014 as our sample period, which includes the financial crisis of 2007/2008, an event that may be considered a structural break. We compare the dynamic relationships of GDP growth ($\Delta\text{GDP}$), unemployment change ($\Delta\text{UNEP}$), inflation change ($\Delta\text{INF}$), industrial production change ($\Delta\text{PROD}$), and three month T-Bill rates (TB) for the USA and unemployment change ($\Delta\text{UNEP}$), inflation change ($\Delta\text{INF}$), industrial production
change ($\Delta$PROD), and short term interest rates (STIR) for the Eurozone. We discuss the suitability of stationary models for these two data sets later in this section. The data series for the USA and Eurozone data sets are displayed in Table 1 and Table 2 respectively. We collected all the USA series from Bloomberg, whereas the Eurozone series were collected from the OECD and the EUROSTAT databases. Unfortunately, we did not include GDP growth in the Eurozone dataset as we were unable to find a seasonally adjusted monthly series for real GDP.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP growth</td>
<td>Annualised log difference in Real GDP</td>
<td>$\Delta$GDP</td>
</tr>
<tr>
<td>Unemployment change</td>
<td>Annualised log difference in unemployment rate</td>
<td>$\Delta$UNEP</td>
</tr>
<tr>
<td>Inflation change</td>
<td>Annualised log difference in PCE deflator</td>
<td>$\Delta$INF</td>
</tr>
<tr>
<td>Industrial production change</td>
<td>Annualised log difference in IPI</td>
<td>$\Delta$PROD</td>
</tr>
<tr>
<td>T-Bill 3m rate</td>
<td>Secondary market rate of 3m T-bill</td>
<td>TB</td>
</tr>
</tbody>
</table>

Table 2: Eurozone data set (monthly and seasonally adjusted); Sources: EUROSTAT, and OECD.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment change</td>
<td>Annualised log difference in unemployment rate</td>
<td>$\Delta$UNEP</td>
</tr>
<tr>
<td>Inflation change</td>
<td>Annualised log difference in CPI</td>
<td>$\Delta$INF</td>
</tr>
<tr>
<td>Interest rate</td>
<td>Short term interest rate</td>
<td>STIR</td>
</tr>
<tr>
<td>Industrial production change</td>
<td>Annualised log difference in IPI</td>
<td>$\Delta$PROD</td>
</tr>
</tbody>
</table>

4.1 **Business cycle study: Dynamic relationships**

We use the posterior summaries described in Section 3.2 to study the characteristics of the USA and Eurozone business cycles, by focusing on the dynamic relationships between different pairs of macroeconomic variables. We produce heat maps of the conditional distri-
bution of $y_{t,i}$ given $y_{t-1,s}$ for $i = 1, \ldots, p$ and $s = 1, \ldots, p$ for $s \neq i$. The colour represents the conditional density with brighter colours representing higher density values, which are linked to more peaked distributions.

Figures 1, 2, 3, 4, and 5 display the dynamic relationships of the macroeconomic variables of the USA data set. We begin with the lagged relationship between $\Delta$INF and $\Delta$UNEP which is displayed in Figure 1. The left panel displays the conditional density of $\Delta$INF at $t$ given $\Delta$UNEP at $t - 1$. Higher densities values are observed when $\Delta$UNEP at $t - 1$ is between $-0.2$, and $0.2$ and $\Delta$INF at $t$ is between $0.01$, and $0.02$. Expected $\Delta$INF is fairly constant and doesn’t seem to be affected by changes in the lagged values of $\Delta$UNEP. A similar picture emerges when we switch over to $\Delta$UNEP at $t$, given $\Delta$INF at $t - 1$ (shown in the right panel), expected $\Delta$UNEP doesn’t seem to be affect by different lagged values of $\Delta$INF. A closer look shows that this lagged relationship is slightly U-shaped with higher expected unemployment when inflation is decreasing or rapidly increasing and a constant level of unemployment when $\Delta$INF is between $0.01$. These findings are contrary to the traditional theory of the Philips Curve which posits an inverse relationship between inflation and unemployment but is more in line with Friedman (1968), and Phelps (1970) who argue that in the short-run when inflation is high, unemployment is low, but in the long run this is not the case. This view is also supported by theories on the expectation-augmented Philips curve and the new-Keynesian Philips curve, both of which imply that increased inflation can lower unemployment but only temporarily, see Clarida et al. (1999), and Blanchard and Gali (2007). As Friedman (1968) and Mankiw (2001) argue we are far from being certain of what their dynamic relationship really is.

Figure 2 and Figure 3 display the lagged relationship between $\Delta$UNEP and $\Delta$PROD and the lagged relationship between $\Delta$UNEP and $\Delta$GDP, respectively. We have grouped them together because the dynamic relationships we observe are very similar. Let’s focus first on $\Delta$UNEP at $t$ and how it is affected by values of $\Delta$PROD and $\Delta$GDP at $t - 1$. For high negative values of $\Delta$PROD at $t - 1$, expected $\Delta$UNEP is high, it decreases sharply until $\Delta$PROD
is 0, and then is stable around $-0.05$. The relationship of $\Delta \text{UNEP}$ at $t$, given $\Delta \text{GDP}$ at $t - 1$ is the same; it sharply decreases when $\Delta \text{GDP}$ starts to increase and then it stabilises around $-0.05$. Looking at the flip side of the dynamic relationship, we observe that when the values of $\Delta \text{UNEP}$ at $t - 1$ are in $(-0.2, 0.1)$, expected $\Delta \text{PROD}$ and $\Delta \text{GDP}$ are stable around 0.03 and 0.025 respectively. So in periods of low unemployment, both productivity and output growth are expected to be stable. When $\Delta \text{UNEP}$ begins to increase, from lagged values greater than 0.1, both expected $\Delta \text{productivity}$ and $\Delta \text{GDP}$ are sharply decreasing. Our results follow Okun’s law (Okun, 1962) which states that production and output decrease as unemployment increases. Our results are consistent with Abel and Bernanke (2005) who find that the magnitude of the decrease declines. In addition we observe that in periods of high productivity and output growth, there is little change in expected unemployment.

The dynamic relationship between $\Delta \text{INF}$ and $\Delta \text{PROD}$ is displayed in Figure 4. The conditional distribution of $\Delta \text{INF}$ at time $t$, given $\Delta \text{PROD}$ at $t - 1$, shown on the left panel, is more peaked when lagged values of $\Delta \text{PROD}$ are between 0 and 0.05. Expected $\Delta \text{INF}$ seems to decrease when values of $\Delta \text{PROD}$ are increasing. When the variables are switched around, and we look at different lagged values of $\Delta \text{INF}$ and how they effect $\Delta \text{PROD}$ at $t$, we observe a slow decreasing linear trend when lagged $\Delta \text{INF}$ values are increasing. Based on these results the lagged relationship between inflation growth and industrial production growth is an inverse linear one. When inflation is high and rising, industrial production slows down, as it becomes costly to acquire materials and services due to the rising prices. There is also a link with high borrowing costs, as high inflation leads to contractionary monetary measures, such as increases in interest rate, which make it costly for firms to borrow and invest in production, see Gordon (2004) and Miles et al. (2012).

Figure 5 displays the lagged relationship between $\Delta \text{GDP}$ and TB. In the left panel we have the conditional distribution of $\Delta \text{GDP}$ at $t$ given TB at $t - 1$, which is peaked when the lagged values of the TB rate are around 1%. The variables have a threshold relationship. When lagged values of TB are increasing (from 0 to 1.5%), expected $\Delta \text{GDP}$ increases, but it
levels off around 0.025 when lagged values of TB pass the 2% mark. The conditional distribution of TB at \( t \) given \( \Delta GDP \) at \( t - 1 \) is displayed in the right panel. It is more peaked when lagged values of \( \Delta GDP \) are between 0 and 0.05 with TB at \( t \) remaining around 1% regardless of the value of lagged \( \Delta GDP \). In summary, we can state that when real interest rates are low, GDP growth increases, possible due to lower borrowing costs, leading firms to invest more. However, if interest rates are high GDP growth slows down, without any sharp decreases. One possible explanation of these observations, is inflation targeting, which started in the US in the early 1990’s. Bernanke et al. (2001), and Mishkin and Schmidt-Hebbel (2007), find that inflation targeting leads to lower interest rates and reduces inflation and output volatility.

The dynamic relationships between the macroeconomic variables of the Eurozone data set, displayed in Figures 6, 7, 8, and 9, are different to those of the USA data. We have observed more unstable, nonlinear relationships, with threshold effects.

In Figure 6 we look at the lagged relationship between \( \Delta INF \) at \( t \) given \( \Delta UNEP \) at \( t - 1 \) shown in the left panel, and in the right panel we switch the variables around and look at the lagged relationship between \( \Delta UNEP \) at \( t \) given \( \Delta INF \) at \( t - 1 \). When \( \Delta UNEP \) in the previous lag is in the interval \((-0.1, 0.1)\), current \( \Delta INF \) remains unchanged around 0. However, for higher lagged values of \( \Delta UNEP \), in the interval \((0.15, 0.25)\), current \( \Delta INF \) begins to increase. What we have observed is a threshold relationship, with 0.1 being the \( \Delta UNEP \) value that triggers higher current inflation. When we flip the variables, at first glance we fall into the trap of thinking that current \( \Delta UNEP \) is not really affected by the previous value of \( \Delta INF \). A closer look, however, reveals that when the lagged values of \( \Delta INF \) are between \(-0.15 \) and \(-0.05\), \( \Delta UNEP \) decreases, then stabilises before it starts to slowly decrease again when lagged \( \Delta INF \) values are between 0.05 and 0.1. Like the USA data, we cannot confirm that we have explained the dynamic trade-off between unemployment and inflation. We do however observe an inverse relationship, with a possible threshold effect. Once again the statements of Friedman (1968) and Phelps (1970) that the relationship between unem-
ployment and inflation may not be negative are supported. Some levels of unemployment may not have an effect on inflation and when unemployment is high it does not mean that inflation is low.

Figure 7 displays the lagged relationship between $\Delta \text{PROD}$ at $t$ given $\Delta \text{UNEP}$ at $t-1$ in the left panel, and the lagged relationship between $\Delta \text{UNEP}$ at $t$ given $\Delta \text{PROD}$ at $t-1$ in the right panel. The relationship between these two variables is different to the one we have observed for the USA data, see Figure 2. The conditional density can be best described as a step function. When the lagged values of $\Delta \text{UNEP}$ are in the interval $(-0.15, 0.10)$, current $\Delta \text{PROD}$ is around 0.05, it remains constant at 0 when lagged values of $\Delta \text{UNEP}$ are in $(-0.05, 0.15)$, and it starts to fall rapidly when $\Delta \text{UNEP}$ lagged values are in $(0.15, 0.25)$. This implies that in periods when unemployment is low, industrial production is constant, but when unemployment is high it sharply declines. Looking at the flipped relationship, we observe that when the previous lags’ values of $\Delta \text{PROD}$ are in $(-0.25, -0.15)$, current $\Delta \text{UNEP}$ is between 0.15 and 0.2, it shifts up to 0.25 when previous $\Delta \text{PROD}$ is in $(-0.1, -0.25)$, and then quickly drops between 0.05 and $-0.1$ when previous $\Delta \text{PROD}$ is in $(-0.05, 0.1)$. In other words when industrial production is low, expected unemployment rate is high, and it drops when industrial production is high. These results are contrary to the search theories of Lucas and Prescott (1974) and Pissarides (1990), which imply that an increase in growth leads to an increased rate of job turnover, that results in a higher rate of unemployment. However, Aghion and Howitt (1994) states that growth has two effects on unemployment, the capitalisation effect and the creative destruction effect. Under the capitalisation effect, an increase in growth leads to increased returns for firm creation and an increasing number of firms entering the market. Under the creative destruction effect, an increase in growth reduces the duration of job matching, which in turn raises the equilibrium level of unemployment by raising job separation rate and discouraging the creation of job vacancies, thus reducing the job finding rate.

In Figure 8 we look at the lagged relationship between $\text{STIR}$ at $t$ given $\Delta \text{UNEP}$ at $t-1$
shown in the left panel, and in the right panel we switch the variables around and look at the lagged relationship between $\Delta\text{UNEP}$ at $t$ given STIR at $t-1$. It appears that previous values of $\Delta\text{UNEP}$ have no effect on the current short term interest level. This can be attributed to the fact that the European central bank kept nominal interest rates low since 2008, to encourage growth. The relationship between lagged values of STIR and current $\Delta\text{UNEP}$ is definitely not constant and not linear. When previous values of STIR are in $(0, 0.5)$, current $\Delta\text{UNEP}$ is in $(-0.1, 0.1)$, it shoots up to $(0.15, 0.20)$ when previous STIR is close to 1%, levels off around 0 when previous STIR values are in $(1.5, 2.5)$, starts to drop to $-0.05$ when STIR values are in $(3, 3.5)$, levels off again around 0.055 and when previous STIR values increase beyond 4.5%, current $\Delta\text{UNEP}$ sharply increases. In summary, low values of previous short term interest rates have little effect on current unemployment rate, whereas the high values at first instance could lead to a drop in current unemployment rate followed by a sharp increase in current unemployment, when some threshold is reached, (in this Eurozone sample the threshold is 4.5% ).

The lagged relationship between $\Delta\text{INF}$ and $\Delta\text{PROD}$ is displayed in Figure 9. The conditional distribution of $\Delta\text{PROD}$ at $t$ given $\Delta\text{INF}$, which is displayed in the left panel, shows that lagged values of $\Delta\text{INF}$ do not alter the expected level of $\Delta\text{PROD}$, which appears constant around 0.025. When the variables are switched over, and we consider the conditional distribution of $\Delta\text{INF}$ at $t$ given $\Delta\text{PROD}$ at $t-1$, a different picture emerges. The lagged relationship is nonlinear, as expected $\Delta\text{INF}$ begins to decrease when lagged values of $\Delta\text{PROD}$ are increasing, with the threshold point around $-0.05$. Again, these results could be linked to the effects of inflation targeting. It encourages lower inflation, and lower output volatility.
4.2 Impulse response functions

The standard tool for tracking the response of macroeconomic variables to fiscal or monetary shocks is the impulse response function. Recall from Section 3.2, that we follow the approach of Koop et al. (1996), and construct generalised impulse response functions. We
use the following function,

$$G_{t}(n, \nu_t, \mathcal{F}_{t-1}) = E[y_{t+n}|\nu_t, \mathcal{F}_{t-1}] - \mu$$

where $\nu_t$ is the shock at time $t$, $\mathcal{F}_{t-1}$ is past information up to time $t - 1$, and $\mu = E(y_t)$, the long-run marginal mean of $y_t$. 
Figure 7: Eurozone data: $\Delta$PROD at $t$ given $\Delta$UNEP at $t - 1$ (left), and $\Delta$UNEP at $t$ given $\Delta$PROD at $t - 1$ (right).

Figure 8: Eurozone data: STIR at $t$ given $\Delta$UNEP at $t - 1$ (left), and $\Delta$UNEP at $t$ given STIR at $t - 1$ (right).

Figure 9: Eurozone data: $\Delta$INF at $t$ given $\Delta$PROD at $t - 1$ (left), and $\Delta$PROD at $t$ given $\Delta$INF at $t - 1$ (right).

We study different negative shocks to interest rates, for both the US and Eurozone datasets. For the US we focused on shocks for the following two months: November 2006 when the T-Bill 3m rate was at its peak, and June 2007 when, after six consecutive months of decreases in the rate, it started increasing again. These months fall in the period close to and during
the burst of the US housing bubble. In terms of the size of the shock we started with a drop of 1.5% from November 2006 onwards, and then halved the shock to see if there would be a change. Again from June 2007 onwards we started with a drop of 0.46%, and then halved the shock to see if there would be a change in the effect on the other variables. Solid lines represent the big shock and dashed lines the small (halved) shock. Figure 10 displays the impulse response functions for the US monetary shocks. The November 2006 shocks are displayed on the left panel. The effect of both shocks on unemployment is nonlinear, there is an increase in unemployment at a faster rate than the rate of decrease that follows. The only difference between the large and smaller shock is the sharpness of the increase. The size of the shock is irrelevant when it comes to inflation growth, industrial production growth and output growth. The effect on inflation is linear, with inflation growth decreasing at a slow rate. For industrial production the effect is nonlinear, the rate of decrease is higher at first and then it slows down, whereas for output growth there is a very small effect, as it decreases for a short period and then it levels off. With the exception of the effect on inflation growth, which is the same as the shocks of November 2006, the effects of the June 2007 on unemployment growth, industrial production growth, and output growth are slightly different. The effect of the large shock on unemployment growth is negative and linear, with the a slow rate of decay. However the effect of the smaller shock is nonlinear, the rate of decay is larger at first and then becomes smaller. The two shocks have the same effect on industrial production and output growth.

For the Eurozone we focused on shocks for one month: November 2007, the time that interest rates were at their peak. We considered a negative shock of 2.6% from November 2007 onwards, and smaller shock of 1.3%. The results are displayed in Figure 11, with solid lines representing the big shock and dashed lines the small (halved) shock. Both negative shocks led to a fall in \( \Delta \text{UNEP} \) and \( \Delta \text{INF} \) and an increase in \( \Delta \text{PROD} \). The smaller shock is slightly more persistent than the larger shock, as it’s effect decays less quickly in the case of \( \Delta \text{UNEP} \) and \( \Delta \text{INF} \). In terms of \( \Delta \text{PROD} \) there is only a marginal difference in shock
persistence, and the positive effect of the negative interest rate shock on $\Delta \text{PROD}$ is non-linear, a result consistent with the findings of Adda and Scoru (1997), Stock and Watson (1999), Eggerston and Woodford (2003), and Hansen and Senhadji (2013).

Figure 11: Eurozone data set: November 2007 shock to interest rates, red $- \Delta \text{UNEP}$, cyan $- \Delta \text{INF}$, and green $- \Delta \text{PROD}$
4.3 Out-of-sample predictive performance

We compare the out-of-sample predictive performance of the Bayesian nonparametric VAR model with other Bayesian VAR specifications, a stationary linear model, namely the BVAR with the independent Normal-Wishart prior, and a non-stationary non-linear model, namely the TVP-VAR with stochastic volatility (TVP-SV-VAR) of Primiceri (2005). The comparison metric is the log-predictive score calculated as,

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

where \(T\) is the size of the time series, \(s\) is the time from where the prediction starts, and \(h\) is the predictive horizon. We looked at \(h = 1, 2,\) and 4 months. Smaller scores indicate better predictive performance. We calculate the log-predictive scores for all the variables jointly, and for each variable marginally. For the stationary models we also consider different lags, and compare the BVAR with 1, 2, 3, and 4 lags, the BayesNP-VAR with 1, 2, and 3 lags to the non-stationary TVP-SV-VAR with 1 lag.

Table 3 displays the log-predictive scores for the USA data, and Table 4 displays the log-predictive scores for the Eurozone data, with the lowest scores in bold. For the USA data, the BayesNP-VAR specification has the best overall predictive performance as well as the best predictive performance when focusing on each variable. The differences in log-scores between the different lags of the BayesNP-VAR model are marginal. The BVAR specification doesn’t perform well at any lag, leaving the TVP-SV-VAR as the closest competitor. The same is true for the Eurozone data, with the only difference being the log-predictive scores for the short term interest rate, where TVP-SV-VAR performed marginally better for \(h = 1\) and \(h = 2\).

To provide a quick and clear reference for the good predictive performance of the BayesNP-VAR and how it compares with the other models we looked at the difference in log-scores between each model and the BVAR(1). Figure 12 displays this difference. The lower its position on the plot the better the model’s predictive performance. Each model has its own
Table 3: Comparison of log-predictive scores of BVAR, TVP-SV-VAR, and BayesNP-VAR model specifications over three predictive horizons (h = 1, h = 2, and h = 4) for the USA data set.

<table>
<thead>
<tr>
<th>Model</th>
<th>$h = 1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>$h = 2$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>$h = 4$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>$\Delta$GDP</td>
<td>$\Delta$UNEP</td>
<td>$\Delta$INF</td>
<td>$\Delta$PROD</td>
<td>TB</td>
<td>Overall</td>
<td>$\Delta$GDP</td>
<td>$\Delta$UNEP</td>
<td>$\Delta$INF</td>
<td>$\Delta$PROD</td>
<td>TB</td>
<td>Overall</td>
<td>$\Delta$GDP</td>
</tr>
<tr>
<td>BVAR(3)</td>
<td>-338</td>
<td>-82</td>
<td>-74</td>
<td>-83</td>
<td>-82</td>
<td>-18</td>
<td>-121</td>
<td>-74</td>
<td>-49</td>
<td>-55</td>
<td>-62</td>
<td>99</td>
<td>390</td>
<td>-63</td>
</tr>
</tbody>
</table>
Table 4: Comparison of log-predictive scores of BVAR, TVP-SV-VAR, and BayesNP-VAR model specifications over three predictive horizons
\((h = 1, h = 2, \text{ and } h = 4)\) for the Eurozone data set.

<table>
<thead>
<tr>
<th>Model</th>
<th>(h = 1)</th>
<th></th>
<th></th>
<th></th>
<th>(h = 2)</th>
<th></th>
<th></th>
<th>(h = 4)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>(\Delta\text{UNEP})</td>
<td>(\Delta\text{INF})</td>
<td>(\Delta\text{STIR})</td>
<td>(\Delta\text{PROD})</td>
<td>Overall</td>
<td>(\Delta\text{UNEP})</td>
<td>(\Delta\text{INF})</td>
<td>(\Delta\text{STIR})</td>
<td>(\Delta\text{PROD})</td>
<td>Overall</td>
</tr>
<tr>
<td>BVAR(2)</td>
<td>-279</td>
<td>-79</td>
<td>-78</td>
<td>-42</td>
<td>-79</td>
<td>-167</td>
<td>-71</td>
<td>-61</td>
<td>1</td>
<td>-71</td>
<td>96</td>
</tr>
<tr>
<td>BVAR(3)</td>
<td>-276</td>
<td>-79</td>
<td>-77</td>
<td>-42</td>
<td>-78</td>
<td>-159</td>
<td>-71</td>
<td>-62</td>
<td>0</td>
<td>-71</td>
<td>172</td>
</tr>
<tr>
<td>BVAR(4)</td>
<td>-270</td>
<td>-78</td>
<td>-76</td>
<td>-41</td>
<td>-77</td>
<td>-155</td>
<td>-70</td>
<td>-60</td>
<td>7</td>
<td>-70</td>
<td>236</td>
</tr>
<tr>
<td>BayesNP-VAR(1)</td>
<td>-463</td>
<td>-139</td>
<td>-124</td>
<td>-56</td>
<td>-144</td>
<td>-354</td>
<td>-111</td>
<td>-99</td>
<td>-29</td>
<td>-130</td>
<td>-199</td>
</tr>
<tr>
<td>BayesNP-VAR(2)</td>
<td>-463</td>
<td>-139</td>
<td>-125</td>
<td>-54</td>
<td>-147</td>
<td>-359</td>
<td>-112</td>
<td>-100</td>
<td>-28</td>
<td>-133</td>
<td>-206</td>
</tr>
</tbody>
</table>
symbol and colour: + - BVAR(2) and ○ - BVAR(3) are in green, ◊ - BayesNP-VAR(1), ▽ - BayesNP-VAR(2), and △ - BayesNP-VAR(3) are in blue, and □ - TVP-SV-VAR(1) is in red.

Figure 12: Differences in log-predictive scores: BVAR(1) is the baseline model, each model has its own symbol. + - BVAR(2) and ○ - BVAR(3) are in green, ◊ - BayesNP-VAR(1), ▽ - BayesNP-VAR(2), and △ - BayesNP-VAR(3) are in blue, and □ - TVP-SV-VAR(1) is in red. The x-axis labels joint and marginal performance for each predictive horizon.

Our results show that the linear stationary BVAR, regardless of the number of lags we have considered, has poor predictive performance for both the USA and Eurozone data. Both BayesNP-VAR and TVP-SV-VAR models account for nonlinearity in the conditional mean, and heteroscedasticity in the conditional variance, and hence, are better suited to capturing the nonlinear dynamic relationships between variables, leading to better predictive performance. The difference between these two models is the stationarity assumption, the BayesNP-VAR is stationary, whereas the TVP-SV-VAR is not. Our motivation for choosing the USA and the Eurozone data sets, was to see how these economies behave, and see if their dynamics are better described by a stationary model or a non-stationary model. Our results demonstrate that the BayesNP-VAR model is better suited to both data sets for the sample period of March 1995 to December 2014. To understand the relevance of the stationarity assumption for out-of-sample predictive performance we decided to produce plots of the cumulative log-predictive scores to compare these two models more closely. The cumu-
lative log-predictive scores were calculated as,

\[- \sum_{i=s}^{s+k} \log p(y_{i+h} | y_1, \ldots, y_i), \quad \text{for} \quad k = 0, 1, \ldots, T - h - s\]

where $T$ is the size of the time series, $s$ is the time from where the prediction starts (the starting point is December 2010), and $h$ is the predictive horizon. Again we considered predictive horizons of $h = 1, 2,$ and $4$ months. Figures 13 and 14 display the cumulative log-predictive scores, over the course of the prediction period, for the USA and Eurozone data sets respectively. The cumulative log-predictive scores of the BayesNP-VAR(1) model are represented by the solid line and the cumulative log-predictive scores of the TVP-SV-VAR(1) by the dashed line. The BayesNP-VAR(1) outperforms the TVP-SV-VAR(1) for all predictive horizons, in both data sets. Therefore in terms of short term forecasting the stationary BayesNP-VAR model is the best model for both economies. The gap between the two lines is narrow for the USA data, and wide for the Eurozone data, and as we move to longer predictive horizons (from $h = 1$ to $h = 2$ and to $h = 4$) the gap widens even more.

Figure 13: Comparison of log-predictive scores for USA data set. BayesNP-VAR(1) – solid line, TVP-SV-VAR(1) – dashed line.
$h = 1$  \hspace{2cm}  $h = 2$  \hspace{2cm}  $h = 4$

Figure 14: Comparison of log-predictive scores for Eurozone data set. BayesNP-VAR(1) – solid line, TVP-SV-VAR(1) – dashed line.

5 Discussion

This paper introduces a new approach to VAR modelling. 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 departures from linearity and normality in both the conditional mean and variance. Our empirical results, for both the USA and Eurozone data, indicate that some macroeconomic variables are nonlinearly related, including threshold type effects, and U-shaped relationships. These findings are consistent with modern economic theory. In terms of out of sample predictions the BayesNP-VAR outperforms all models for all horizons for both the USA and the Eurozone datasets. However, our model assumes stationarity, 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.
References


### 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 renormalized stick-breaking construction of Griffin (2014) with $K$ atoms as our truncation. This implies a transition density of the form

$$p(y_t | y_{(t-L):(t-1)}) = \frac{\sum_{j=1}^{\infty} w_j k(y_{(t-L):(t-1)} | \theta_j)}{\sum_{j=1}^{\infty} w_j k(y_{(t-L):(t-1)} | \theta_j)}$$
where \( w_j = V_j \prod_{m<j} (1 - V_m) \) and \( V_k \overset{i.i.d.}{\sim} \text{Be}(1, M) \) for \( 1 \leq k \leq K \) and \( \theta_j \overset{i.i.d.}{\sim} H \).

**Updating \( \mu \)**

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

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

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

**Updating \( \xi \)**

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

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

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 $\rho$**

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

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

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 $\rho^*$**

The full conditional density of $\rho^*_{j,i}$ is proportional to

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

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 $\Lambda$**

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

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

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 $\phi$**

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

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

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

The full conditional distribution for $\delta_{h, 2 \leq k \leq q}$ 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 $S_{\mu}$

The full conditional distribution of $S_{\mu,i}$ 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

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

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 $\text{Ga} \left( 1 + K, 1 - \sum_{i=1}^K \log(1 - V_i) \right)$.