ON THE HIERARCHICAL SPECIFICATION, ESTIMATION AND INFERENCE OF NON-PARAMETRIC VECTOR AUTOREGRESSIVE MODELS WITH DYNAMIC FACTOR

PARUSH ARORA[†]

Abstract. This paper introduces a non-parametric vector autoregressive model with dynamic factor (NPVAR-DF) through a hierarchical Bayesian approach. A framework is presented that aims at effectively capturing dynamic relationships among variables and enabling the incorporation of extensive information sets. The paper considers the specification, identification and estimation of the NPVAR-DF model, allowing the model to be efficiently fit by MCMC algorithms. Issues related to model comparison and extensions to settings with autocorrelated errors and qualitative variables are also considered. The NPVAR-DF model successfully identifies non-linear associations between macroeconomic variables by employing post-war US data. The dynamic factor captures the business cycle component, which aligns with officially declared recession periods.

Keywords: Bayesian inference; NPVAR-DF model; Markov chain Monte Carlo; Business Cycle; hierarchical framework.
JEL Codes: C11, C14, C15, C32, C52, E31, E32, E37, E43, E47.

1. Introduction

Vector Auto-Regressive (VAR) models have significantly contributed to empirical macroeconomics literature, following the seminal work of Sims et al. (1986). Its linear form is intuitive and captures the dynamics inter-dependency of multiple macroeconomic variables. Its popularity stems from its broad applicability, encompassing techniques such as time series modelling and forecasting. Let $y_t = \{y_{1t}, y_{2t}, \ldots, y_{Qt}\}'$ be a Q dimensional vector for all $t = 1, \ldots, T$. For a given t, the basic VAR model associates each variable with their own and other variable's lag values as

$$y_t = c + \sum_{p=1}^{P} B_p y_{t-p} + \epsilon_t. \tag{1}$$

where c is a $Q \times 1$ dimensional vector of intercepts, B_p is $Q \times Q$ dimensional matrix of parameters for y_{t-p} with $p = 1, \ldots, P$ being the number of lags. It assumes $E(\epsilon_t) = 0$ and $var(\epsilon_t) = \Omega$ with Ω being a $Q \times Q$ dimensional variance covariance matrix.

Beyond its basic form, the VAR model can capture complicated relations between the variables. For instance, following Hamilton (1989), a lot of research has been done on regime-switching models (Hansen (1992), Chib (1996), Vigfusson (1997), Kim and Nelson (1998), Kim et al. (2005), Sims and Zha (2006)). Tong (1978) explored structural instability in univariate models and Balke (2000), Huang et al. (2005) and, Van Robays (2016) furthered it to VAR models. By relaxing the assumption

[†] Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697. E-mail addresses: parusha@uci.edu.

of the static parameter value, time-varying parameter VAR (TVP-VAR) models allow the elements in $\{B_j\}_{j=1}^{p}$ to vary (Canova (1993), Cogley and Sargent (2005), Primiceri (2005), Chan and Jeliazkov (2009) and many others).

One theme common to many extensions, including those stated above, is the assumption of linearity – allowing the model to be tractable in terms of estimation and interpretation. However, the linear model is often unable to explain all the macroeconomic relations, especially when the non-linearity is prominent (Engle (1982), Robinson (1983) and Vieu (1995)). In response to a growing need, many non-parametric time series models were introduced. These include but are not limited to the threshold autoregressive (TAR) model by Tong and Lim (1980) and the exponential autoregressive model by Haggan and Ozaki (1981). Following in the footsteps, the assumption of non-linearity was extended to the VAR models by Härdle et al. (1998b), who estimated the non-parametric version using polynomial fitting. Hastie and Tibshirani (1990) introduced generalized additive models, which gave structure to equations having multiple non-linear functions. Jeliazkov (2013) used this generalized additive model and estimated the non-parametric VAR (NPVAR) model under the Bayesian hierarchical framework. Kalli and Griffin (2018) estimated the non-parametric VAR model by modelling the stationary and transition densities using Bayesian non-parametric methods.

Another limitation of the VAR model is its inability to incorporate many variables due to the dynamic inter-dependency structure. Every extra variable introduced in a VAR model consumes $2qp + 1$ degrees of freedom. To deal with it, Bernanke et al. (2005) introduced the FAVAR model, which compresses the information from a large number of variables into a few factors, thus not steeply penalizing the degrees of freedom. Bai et al. (2016) extended the FAVAR literature by studying the identification restriction and proposing a two-step sequential estimation technique. Amir-Ahmadi and Uhlig (2009) estimated the FAVAR model using the MCMC methods under the Bayesian likelihood-based approach. Another innovation explored multivariate time series models with dynamic factors (Geweke (1977), Sargent et al. (1977). Stock and Watson (2011) summarizes the three major categories of estimation approaches used in the literature for Dynamic factor models.

Though the extension of the VAR model allows for non-linearity and factors, no attempt has been made to include both in the same model. One of the reasons could be that the model estimation becomes more involved when multiple extensions are incorporated. This paper proposes a Bayesian hierarchical framework through which the VAR model with dynamic factor can be estimated under the non-parametric framework. In other words, the paper makes the first attempt to estimate the non-parametric VAR model with dynamic factors and thus calls it the NPVAR-DF model.

The remainder of the paper is structured as follows. Section 2 discusses the hierarchical framework of the NPVAR-DF model. Section 3 presents the prior distributions used for each parameter in the model. Section 4 discusses the identification restrictions required for estimating loadings, factors and unknown additive functions. Section 5 lays out the estimation procedure, which uses an efficient fitting algorithm based on MCMC simulation techniques. Section 6 explains the model comparison procedure. Section 7 suggests some of the model extensions, including serially correlated error and the inclusion of qualitative variables. Section 8 considers applying the NPVAR-DF model to the post-war US economic data and making inferences from the results. Section 9 concludes this article.

2. Hierarchical Model Framework for NPVAR model

The paper proposes a hierarchical structure for the dynamic system of Q regression equations under the Gaussian state space framework. Let the q^{th} equation is modelled through the additive form as in Hastie and Tibshirani (1990). For all $t = 1, 2, \dots, T$, $p = 1, 2, \dots, P$ and $q = 1, \dots, Q$

$$\begin{aligned} y_{1t} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{1qpt}(y_{q,t-p}) + a_1' f_t + \epsilon_{1t} \\ &\vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\ y_{Qt} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{Qqpt}(y_{q,t-p}) + a_Q' f_t + \epsilon_{Qt} \end{aligned} \tag{2}$$

where y_{qt} is the dependent variable indexed by q, representing the equation it belongs to at time t and g_{rqpt} is a function in equation r evaluated at the q^{th} variable at time t with lag p, where $r = 1, \dots, Q$. The factor f_t is a $D \times 1$ dimensional vector which enters the model only as a covariate along with a_q, which is a $D \times 1$ vector of factor loadings. The paper imposes the Gaussian distribution on the error term where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{Qt})' \sim N(0, \Omega_1)$ with

$$\Omega_1 = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1Q} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{Q1} & \omega_{Q2} & \cdots & \omega_{QQ} \end{bmatrix}.$$

The factors follow an AR(1) process, which is described using the transition equation as follows

$$f_t = F f_{t-1} + Z_t \delta + v_t. \tag{3}$$

The covariate Z_t is a $D \times L$ matrix which is structured as

$$\begin{aligned} Z_t &= \begin{bmatrix} z_{1t}' & 0 & \cdots & 0 \\ 0 & z_{2t}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_{Dt}' \end{bmatrix} \\ &= \begin{bmatrix} \{z_{11t}, \dots, z_{1D_1t}\} & 0 & \cdots & 0 \\ 0 & \{z_{21t}, \dots, z_{2D_2t}\} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \{z_{D1t}, \dots, z_{DD_Dt}\} \end{bmatrix} \end{aligned}$$

where z_{dt} is $D_d \times 1$ vector for all $d = 1, 2, \dots, D$ with $\sum_{d=1}^{D} D_d = L$ and δ is $L \times 1$ vector of slope parameters. The error term v_t follows a Gaussian distribution given as $v_t \sim N(0, \Omega_2)$ where $\Omega_2 = diag\{\sigma_1^2, \dots, \sigma_D^2\}$ is a $D \times D$ variance covariance matrix. The model parameter $F = diag\{\gamma_1, \dots, \gamma_D\}$ is a $D \times D$ matrix containing coefficients for the lag factor terms.

The paper rewrites the model using Jeliazkov (2013) notations to simplify further and facilitate estimation. Let $j = qp$ be the new index for lagged variables on the right-hand side of Eq. 2 and $j = 1, \ldots, J$ where $J = QP$. The functions $g(.)$ are now evaluated at the new notation s_{qjt} representing $y_{q,t-p}$. The q^{th} equation for a given period t in Eq. 2 will look like

$$y_{qt} = g_{q1}(s_{q1t}) + \cdots + g_{qJ}(s_{qJt}) + a_q^{'} f_t + \epsilon_{qt}. \tag{4}$$

Equation 4 is stacked over time and written in the matrix form to facilitate the joint sampling of functions and parameters from their posterior distribution. Let $y_q = \{y_{q1}, y_{q2}, \ldots, y_{qT}\}'$ and $e_q = \{e_{q1}, e_{q2}, \ldots, e_{qT}\}'$. For each function in Eq. 4, let the T observations in the covariate vectors $s_{qj} = \{s_{qj1}, \ldots, s_{qjT}\}'$ determine the corresponding $m_j \times 1$ design point vector $v_{qj} = \{v_{qj1}, \ldots, v_{qjm_j}\}'$ with entries equal the unique ordered values of s_{qj}, that is $v_{qj1} < \ldots < v_{qjm_j}$. Let the corresponding function evaluation vector be denoted by $g_{qj} = \{g_{qj}(v_{qj1}), \ldots\ , g_{qj}(v_{qjm_j})\}'$. The matrix form for equation q in Eq. 2 can be written as

$$y_q = Q_{q1}g_{q1} + Q_{q2}g_{q2} + \cdots + Q_{qJ}g_{qJ} + A_q f + \epsilon_q \tag{5}$$

where $A_q = I_T \otimes a_q'$ is a $T \times DT$ dimensional matrix, $\otimes$ denotes the Kronecker product and $f = \{f_1, \ldots, f_T\}'$ is a $DT \times 1$ vector. The incidence matrix Q_{qj} is $T \times m_j$ dimensional which establishes the relationship between s_{qj} and v_{qj} with entries $Q_{qj}(t, m) = 1$ if $s_{qjt} = v_{qjm}$ and 0 otherwise where $m = 1, \ldots, m_j$. As there may be repeating values in s_{qj}, $m_j \leq T$.

The paper attempts to motivate the non-parametric functions evaluations to be perceived as realizations from a stochastic process, which tries to capture the degree of local variation between neighbouring points. A smoothness hyper-parameter is introduced in the prior of the non-linear functions, which completes the hierarchical structure of the NPVAR-DF model. The vector of function evaluations g_{qj} conditional on the smoothness parameter, τ^2, follows a Gaussian distribution as

$$g_{qj}|\tau_{qj}^2 \sim N(g_{qj0}, \tau_{qj}^2 K_{qj}^{-1}) \tag{6}$$

where K_{qj} is an $m_j \times m_j$ matrix, which will be discussed later. From a Bayesian perspective, τ is interpreted as a smoothness parameter which introduces a penalty to local variation between successive elements of g_{qj} and still allows for any possible value that the elements of g_{qj} can take. From a likelihood-based perspective, it can be interpreted as a roughness penalty term in penalized likelihood estimation (Wahba (1978).

To complete the NPVAR-DF framework, Eq. 3 is stacked over time, and the complete model looks like

$$\begin{aligned} y_1 &= Q_{11}g_{11} + Q_{12}g_{12} \cdots + Q_{1J}g_{1J} + A_1 f + \epsilon_1 \\ &\ \ \vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ y_Q &= Q_{Q1}g_{Q1} + Q_{Q2}g_{Q1} \cdots + Q_{QJ}g_{QJ} + A_Q f + \epsilon_Q \\ H_f f &= Z\delta + v \end{aligned} \tag{7}$$

where

$$H_f = \begin{bmatrix} I_D & 0 & \cdots & 0 \\ -F & I_D & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -F & I_D \end{bmatrix}, \; Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_T \end{bmatrix} \; and \; v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}$$

Following the Gaussian structure, $v \sim N(0, S)$ where $S = diag\{\Omega_S, \Omega_2, \ldots, \Omega_2\}$. The matrix Ω_S is discussed in detail in the next section. Equation 7 will be used to characterize the likelihood function, which will facilitate the calculation of the conditional distribution of non-parametric functions and parameters.

3. Prior Distribution

This section discusses the prior distributions and model representation used in the paper to draw from the posterior distributions of non-parametric functions, factors and other parameters.

3.1. **Prior for g_{qj}.** The non-parametric functions g_{qj} are modelled as the second-order Markov process priors. The prior imposes a structure on the functions where deviations from linearity are penalized. The prior is given as

$$g_{qjm} = \Big(1 + \frac{h_{qjm}}{h_{qj,m-1}}\Big) g_{qj,m-1} - \Big(\frac{h_{qjm}}{h_{qj,m-1}}\Big) g_{qj,m-2} + u_{qjm} \tag{8}$$

where $h_{qjm} = v_{qjm} - v_{qj,m-1}$ (m is the index for unique ordered value) and $u_{qjm} \sim N(0, \tau_{qj}^2 h_{qjm})$. Distribution of the initial state of g_{qj} is necessary for the Markov prior to be proper. The initial state can be modelled as

$$\begin{pmatrix} g_{qj1} \\ g_{qj2} \end{pmatrix} \sim N\left(\begin{pmatrix} g_{qj10} \\ g_{qj20} \end{pmatrix}, \tau_{qj}^2 G_{qj0} \right). \tag{9}$$

The prior for g can be written as $H_{qj} g_{qj} = u_{qj}$ where $u_{qj} \sim N(\hat{\mu}_{qj}, \Sigma_{qj})$ being the error term in the Markov process with $\hat{\mu}_{qj} = (g_{qj10}, g_{qj20}, 0, \ldots, 0)'$ and H_{qj} and Σ_{qj} defined as

$$\begin{aligned} H_{qj} &= \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \frac{h_{qj3}}{h_{qj2}} & -\Big(1 + \frac{h_{qj3}}{h_{qj2}}\Big) & 1 & \ddots & \vdots \\ & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \frac{h_{qjm_j}}{h_{qj(m_j-1)}} & -\Big(1 + \frac{h_{qjm_j}}{h_{qj(m_j-1)}}\Big) & 1 \end{bmatrix} \; and \\ \Sigma_{qj} &= \begin{bmatrix} G_{qj0} & 0 & \cdots & 0 \\ 0 & h_{qj3} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{qjm_j} \end{bmatrix}. \end{aligned}$$

A simple change in variable from u_{qj} to g_{qj} will give the conditional distribution as $g_{qj}|\tau^2_{qj} \sim N(g_{qj0}, \tau^2_{qj}K^{-1}_{qj})$ where the penalty matrix $K_{qj} = H'_{qj}\Sigma^{-1}_{qj}H_{qj}$ and $g_{qj0} = H^{-1}_{qj}\hat{\mu}_{qj}$.

3.2. **Prior for f.** The representation in Eq. 7 is made even more compact by stacking the observation equations over each other to enable the drawing of f from its joint conditional distribution.

$$
\begin{aligned}
&y = Q_1 g_1 + Q_2 g_2 + \cdots + Q_J g_J + Af + \epsilon \\
&H_f f = Z\delta + v
\end{aligned}
\tag{10}
$$

where,

$$
y = \begin{bmatrix} y_1 \\ \vdots \\ y_Q \end{bmatrix}, \; Q_j = \begin{bmatrix} Q_{1j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_{Qj} \end{bmatrix}, \; g_j = \begin{bmatrix} g_{1j} \\ \vdots \\ g_{Qj} \end{bmatrix}, \; A = \begin{bmatrix} A_1 \\ \vdots \\ A_Q \end{bmatrix}, \; e = \begin{bmatrix} e_1 \\ \vdots \\ e_Q \end{bmatrix},
$$

$$
\Omega^*_1 = \begin{bmatrix} I_T \times \omega_{11} & I_T \times \omega_{12} & \cdots & I_T \times \omega_{1Q} \\ I_T \times \omega_{21} & I_T \times \omega_{22} & \cdots & I_T \times \omega_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ I_T \times \omega_{Q1} & I_T \times \omega_{Q2} & \cdots & I_T \times \omega_{QQ} \end{bmatrix},
$$

where $\epsilon \sim N(0, \Omega^*_1)$. A simple change in variable from v to f will lead to $f \sim N(\tilde{f}, K^{-1}_f)$ where $\tilde{f} = H^{-1}_f Z\delta$ and the precision matrix $K_f = H'_f S^{-1} H_f$. Initial states for all the factors are defined as $f_{d1} \sim N(0, \sigma^2_d/(1-\gamma^2_d))$. The initial conditions ensure that the prior distribution is proper. Therefore, the matrix $\Omega_S = diag\{\sigma^2_1/(1-\gamma^2_1), \ldots, \sigma^2_D/(1\gamma^2_D)\}$.

3.3. **Prior for other parameters.** The priors for other coefficients and hyperparameters for $q = 1, \ldots, Q$, $j = 1, \ldots, J$ and $d = 1, \ldots, D$ are given as

$$
\begin{aligned}
\tau^2_{qj} &\sim IG(\frac{\nu_{qj0}}{2}, \frac{V_{qj0}}{2}) \\
\Omega^{-1}_1 &\sim W(r_0, R_0) \\
a &\sim N(a_0, A_0) \\
\gamma_d &\sim TN_{(-1,1)}(\gamma_{d0}, G_{d0}) \\
\sigma^2_d &\sim IG(\frac{\sigma_{d0}}{2}, \frac{S_{d0}}{2}) \\
\delta &\sim N(\delta_0, D_0)
\end{aligned}
\tag{11}
$$

where $a = \{a'_1, \ldots, a'_Q\}'$ is a $QD \times 1$ dimensional vector. The new vector representation of loadings will be useful for specifying its full conditional distribution.

4. Identification Restrictions

Bayesian models with proper priors do not suffer from the problem of identification (Lindley (1971); Poirier (1998)). Due to the additive nature of non-parametric functions and the interaction of factors with loadings, there are two identification issues. The first one is associated with the intercept term as all rows of every incidence matrix sum to 1, leading to the problem of perfect collinearity among g_{qj}. In simple terms, the g_{qj} are correlated by construction since they enter the mean function additively which leads to the emergence of free constants. To see the problem, for any two functions r and s from Eq. 4, the additive form can be written as

$$\begin{aligned} g_{qr} + g_{qs} &= (g_{qr} + \alpha) + (g_{qs} - \alpha) \\ &= g^*_{qr} + g^*_{qs} \end{aligned} \tag{12}$$

where $g^*_{qr} = g_{qr} + \alpha$ and $g^*_{qs} = g_{qs} - \alpha$. Since both forms are observably the same, it is obvious that neither an intercept nor the level of the individual functions is identified. There are multiple ways to anchor the functions, which will lead to their unique estimation. This paper adopts the strategy from Jeliazkov (2013) where the additive functions in q^{th} equation of Eq. 7 are identified by centering the $J-1$ functions as

$$y_q = Q_{q1}g_{q1} + M_0 Q_{q2} g_{q2} + \cdots + M_0 Q_{qJ} g_{qJ} + A_q f + \epsilon_q \tag{13}$$

where M_0 is a $T \times T$ symmetric idempotent mean-differencing matrix defined as

$$M_0 = \left(I_T - \frac{l_T l'_T}{T} \right)$$

Centering allows the first function to capture the intercept of the regression equation. Similarly, Eq. 10 will be updated in the following way

$$y = Q_1 g_1 + M_{00} Q_2 g_2 + \cdots + M_{00} Q_J g_J + Af + \epsilon \tag{14}$$

where,

$$M_{00} = \begin{bmatrix} M_0 & 0 & \cdots & 0 \\ 0 & M_0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & M_0 \end{bmatrix}$$

The second identification problem is associated with the factors and loadings. Since both are unknown, there is a problem related to the scale and sign identification. For any $D \times D$ matrix M_f, the problem can be stated as

$$a'_q f_t = (a'_q M_f)(M_f^{-1} f_t) = a^{*'}_q f^*_t \tag{15}$$

where $a'^*_q = a'_q M_f$ and $f^*_t = M_f^{-1} f_t$. Since both forms are observably identical, it is impossible to identify the scale or sign of the term uniquely. Before introducing the identification restriction, let

$$A_f = \begin{bmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_Q \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1D} \\ a_{21} & a_{22} & \cdots & a_{2D} \\ \vdots & \vdots & & \vdots \\ a_{Q1} & a_{Q2} & \cdots & a_{QD} \end{bmatrix}.$$

There are D^2 free terms in the M_f matrix. The model specifies Ω_2 to be a diagonal matrix (discussed earlier) which imposes $D(D-1)/2$ restrictions on Eq. 3. The identification problem is resolved by fixing the top D^2 part of the matrix A_f to be a lower triangular matrix with diagonal elements restricted to 1 as

$$A_f = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{D1} & a_{D2} & \cdots & 1 \\ a_{(D+1)1} & a_{(D+1)2} & \cdots & a_{(D+1)D} \\ \vdots & \vdots & & \vdots \\ a_{Q1} & a_{Q2} & \cdots & a_{QD} \end{bmatrix}.$$

This further imposes $D^2 - \big(D(D-1)/2\big)$ restrictions, which is necessary and sufficient for estimating the model. The paper adopts these identification restrictions to estimate the model.

5. Estimation

The paper first considers estimating a univariate Gaussian non-parametric model with a single dynamic factor and one lag to motivate the estimation. Let y_t and f_t be scalar quantities. The model will take the following form

$$\begin{aligned} y_t &= g_t(y_{t-1}) + af_t + \epsilon_t \\ f_t &= \gamma f_{t-1} + v_t \end{aligned} \tag{16}$$

The stacked version of this simple univariate model can be written as

$$\begin{aligned} y &= Qg + Af + \epsilon\,, \qquad && \epsilon \sim N(0, \sigma_\epsilon^2 I) \\ H_f f &= v\,, && v \sim N(0, \sigma_v^2 I) \end{aligned} \tag{17}$$

Let θ denote the set of all model parameters. The MCMC estimation for this simple univariate model involves sequential sampling from the conditional distributions. However, sequential sampling of factors and their loadings suffers from the problem of slow convergence and poor mixing. To avoid this, the paper considers sampling factors and their loading jointly through the block sampling scheme suggested in Chib and Jeliazkov (2006). The following steps state the estimation algorithm.

Algorithm 1: MCMC Implementation of the Univariate Gaussian Non-Parametric Dynamic Factor Model

Step 1. Sample $[g|y, \theta_{\{-g\}}] \sim N(\hat{g}, \hat{G})$ where $\hat{G} = (\frac{1}{\tau^2}K + \frac{1}{\sigma_\epsilon^2}Q'Q)^{-1}$ and $\hat{g} = \hat{G}\big(\frac{1}{\tau^2}Kg_0 + \frac{1}{\sigma_\epsilon^2}Q'(y - Af)\big)$ where g_0 is the mean of the prior for g and $K = K_{qj}$ as defined earlier.

Step 2. Sample $[\tau^2|g] \sim IG\Big((\nu_0 + m)/2, \big(V_0 + (g - g_0)'K(g - g_0)\big)/2\Big)$ where ν_0 and V_0 are parameters for the inverse gamma prior for τ^2 and m is the count of unique elements of y.

Step 3. Sample $[\sigma_\epsilon^2|y, \theta_{\{-\sigma_\epsilon^2\}}] \sim IG\Big((r_0 + T)/2, \big(R_0 + (y - Qg - Af)'(y - Qg - Af)\big)/2\Big)$ where r_0 and R_0 are parameters for the inverse gamma prior for σ_ϵ^2.

STEP 4. Sample $[a, f|y, \theta_{-\{a,f\}}]$ in one block as

(1) Sample $[a|y, \theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with tailored proposal $a^* \sim q(\tilde{a}, W)$ and accept a^* with probability

$$\alpha_{MH}(a, a^*) = min\left\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a}, W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a}, W)}\right\}$$

(2) Sample $[f|y, \theta_{-\{f\}}] \sim N(\hat{f}, \hat{F})$ where $\hat{F} = \left(K_f + \frac{1}{\sigma_v^2}A'A\right)^{-1}$ and $\hat{f} = \hat{F}\left(K_f\tilde{f} + \frac{1}{\sigma_v^2}A'(y - Qg)\right)$

STEP 5. Sample $[\gamma|f, \sigma_v^2]$ by MH algorithm with proposal $\gamma^* \sim N(\hat{\gamma}|\hat{G})$ where $\hat{G} = (G_0^{-1} + (f'_{1:T-1}f_{1:T-1})/\sigma_v^2)^{-1}$ and $\hat{\gamma} = \hat{G}(G_0^{-1}\gamma_0 + (f'_{1:T-1}f_{2:T})/\sigma_v^2)$. The proposal γ^* is accepted with the probability

$$\alpha_{MH}(\gamma, \gamma^*) = min\left\{1, \frac{f_N(f_1|0, \sigma_v^2/(1-\gamma^{*2}))}{f_N(f_1|0, \sigma_v^2/(1-\gamma^2))}\right\}$$

STEP 6. Sample $[\sigma_v^2|f, \gamma] \sim IG\Big(\sigma_0 + T/2, \big(S_0 + (f^* - \bar{f})'(f^* - \bar{f})\big)/2\Big)$ where $f^* = (f_1\sqrt{1-\gamma^2}, f_2, \ldots, f_T)$ and $\bar{f} = (0, \gamma f_2, \ldots, \gamma f_T)$.

The tailored proposal density $a^* \sim q(\tilde{a}, W)$ used in Step 4 is generally a multivariate Student's t distribution with low degrees of freedom to ensure heavy tails with $\hat{a}$ and W being the mode and inverse of the negative Hessian at the mode of $[a|y, \theta_{-f}]$ (Chib (1996)). The conditional density of a marginal of f is obtained as

$$\begin{aligned} \pi(a|y, \theta_{\{-f,a\}}) &= \frac{\pi(a|y, \theta_{\{-f\}})\pi(f|y, \theta_{\{-f,a\}})}{\pi(f|y, \theta_{\{-a\}})} \\ \pi(a|y, \theta_{\{-f,a\}}) &\propto \frac{\pi(a|y, \theta_{\{-f\}})}{\pi(f|y, \theta_{\{-a\}})} \end{aligned} \tag{18}$$

which is not difficult to obtain since full conditional densities for a and f are known in the model and $\pi(f|y, \theta_{\{-a\}})$ is absorbed in the constant of proportionality. The sampling of g in step 1 involves the inversion of an $m \times m$ matrix. To avoid inverting using brute force, *Remark 1* explains a trick which leads to computational efficiency.

Remark 1: Sampling of Non-Centered Functions. Since the matrices K and $Q'Q$ are banded, the variance covariance matrix $\hat{G}$ is banded as well. Computational efficiency can be achieved by avoiding the inversion of $(\tau^2 K + \frac{1}{\sigma_\epsilon^2}Q'Q)$. In order to obtain a random draw from $N(\hat{g}, \hat{G})$ efficiently, first sample $u \sim N(0, I)$, and then solve $Pw = u$ for w by back substitution where P is the Cholesky decomposition of $\hat{G^{-1}}$. It follows that $w \sim N(0, \hat{G})$. Adding the mean $\hat{g}$ to w, will be equivalent to drawing $g \sim N(\hat{g}, \hat{G})$. The mean $\hat{g}$ is found by solving $\hat{G}\hat{g} = \frac{1}{\tau^2}Kg_0 + \frac{1}{\sigma_\epsilon^2}Q'(y - Af)$, which is done in $O(T)$ operations by back substitution.

Turning attention to the NPVAR-DF, given the priors for the parameters in Sec. 3 and the identification restrictions in Sec. 4, the MCMC estimation can proceed through an iterative sampling of the following steps.

Algorithm 2: MCMC Implementation of the NPVAR-DF Model

STEP 1. Sample the first function (non-centered) in q^{th} equation as $[g_{q1}|y,\theta_{\{-g_{q1}\}}] \sim N(\hat{g_{q1}},\hat{G_{q1}})$ where

$$\begin{aligned}\hat{G}_{q1} &= \Big(\frac{1}{\tau^2_{q1}}K_{q1} + \frac{1}{\omega_{q|\{-q\}}}Q'_{q1}Q_{q1}\Big)^{-1}\\ \hat{g}_{q1} &= \hat{G}_{q1}\Big(\frac{1}{\tau^2_{q1}}K_{q1}g_{q10} + \frac{1}{\omega_{q|\{-q\}}}Q'_{q1}\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj}g_{qj} - A_q f\big)\Big)\end{aligned}$$

with $\mu_{q|\{-q\}} = E(\epsilon_q|\epsilon_{-q})$ and $\omega_{q|\{-q\}} = Var(\epsilon_q|\epsilon_{-q})$. The centered functions are sampled as $[g_{qj}|y,\theta_{\{-g_{qj}\}}] \sim N(\hat{g_{qj}},\hat{G_{qj}})$ for $j = 2,\ldots,J$ where

$$\begin{aligned}\hat{G}_{qj} &= \Big(\frac{1}{\tau^2_{qj}}K_{qj} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}M_0Q_{qj}\Big)^{-1}\\ \hat{g}_{qj} &= \hat{G}_{qj}\Big(\frac{1}{\tau^2_{qj}}K_{qj}g_{qj0} + \frac{1}{\omega_{q|\{-q\}}}Q'_{qj}M_0\big(y_q - \mu_{q|\{-q\}} - \sum_{r\geq 2,k\neq j} M_0 Q_{qr}g_{qr}\\ &\quad -Q_{q1}g_{q1} - A_q f\big)\Big).\end{aligned}$$

STEP 2. Sample $[\tau^2_{qj}|g_{qj}] \sim IG\Big((\nu_{qj0}+m_j)/2, \big(V_{qj0} + (g_{qj}-g_{qj0})'K_{qj}(g_{qj}-g_{qj0})\big)/2\Big)$.

STEP 3. Sample $[\Omega_1^{-1}|y,\theta_{\{-\Omega_1\}}] \sim W\Big(r_0+T, \big(R_0^{-1} + \sum_{t=1}^T e_t e_t'\big)^{-1}\Big)$ where e_t denotes Q vector of residuals in time period t.

STEP 4. Sample $[a,f|y,\theta_{-\{a,f\}}]$ in one block as

(1) Sample $[a|y,\theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with tailored proposal $a^* \sim q(\tilde{a},W)$ and accept a^* with probability

$$\alpha_{MH}(a,a^*) = min\Bigg\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a},W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a},W)}\Bigg\}$$

(2) Sample $[f|y,\theta_{-\{f\}}] \sim N(\hat{f},\hat{F})$ where

$$\begin{aligned}\hat{F} &= \big(K_f + A'(\Omega_1^*)^{-1}A\big)^{-1}\\ \hat{f} &= \hat{F}\Big(K_f\tilde{f} + A'(\Omega_1^*)^{-1}\big(y - Q_1g_1 - \sum_{j=2}^{J} M_{00}Q_Jg_J\big)\Big)\end{aligned}$$

STEP 5. Sample $[\gamma_d|f_d,\sigma^2_d]$ by MH algorithm with proposal $\gamma^*_d \sim N(\hat{\gamma_d},\hat{G}_d)$ where $\hat{G}_d = (G_{d0}^{-1} + (f'_{\{d,1:T-1\}}f_{\{d,1:T-1\}})/\sigma^2_d)^{-1}$ and $\hat{\gamma}_d = \hat{G}_d\big(G_{d0}^{-1}\gamma_{d0} + (f'_{\{d,1:T-1\}}f_{\{d,2:T\}}/\sigma^2_d)^{-1}\big)$ where $f_{\{d,n:m\}}$ is the vector of d^{th} factor from time n to m. The proposal γ^*_d is accepted with the probability

$$\alpha_{MH}(\gamma_d,\gamma^*_d) = min\Bigg\{1, \frac{\phi(f_1|0,\frac{\sigma^2_d}{1-\gamma^{*2}_d})}{\phi(f_1|0,\frac{\sigma^2_d}{1-\gamma^2_d})}\Bigg\}$$

where $\phi(.)$ is the Gaussian density.

STEP 6. Sample $[\sigma^2_d|f_d,\gamma_d] \sim IG\Big((\sigma_{d0}+T)/2, \big(S_{d0} + (f^*_d - \bar{f}_d)'(f^*_d - \bar{f}_d)\big)/2\Big)$ where $f^*_d = (f_{k1}\sqrt{1-\gamma^2_d}\ , f_{k2},\ldots,f_{kT})'$ and $\bar{f}_d = (0,\gamma_d f_{k2},\ldots,\gamma_d f_{kT})'$.

STEP 7. Sample $[\delta|y,\theta_{\{-\delta\}}] \sim N(\hat{\delta},\hat{D})$ where

$$\begin{aligned}\hat{D} &= \big(D_0^{-1} + Z'S^{-1}Z\big)^{-1}\\ \hat{\delta} &= \hat{D}\big(D_0^{-1}\delta_0 + Z'S^{-1}(H_f f)\big)\end{aligned}$$

A slight change in representation of Eq. 14 is required to obtain $\pi(a|y,\theta_{\{-f\}})$ which is essential to obtain $\pi(a|y,\theta_{\{-f,a\}})$ in Eq. 18.

$$y = Q_1 g_1 + M_{00}Q_2 g_2 + \cdots + M_{00}Q_J g_J + F_f a + \epsilon \tag{19}$$

where $F_f = diag\{\tilde{F},\ldots,\tilde{F}\}$ is a $TQ \times TD$ dimensional vector with $\tilde{F} = \{f_1',\ldots,f_T'\}'$. Thus, the full conditional density of a will take the form $[a|y,\theta_{-\{a\}}] \sim N(\hat{a},\hat{A})$ where

$$\begin{aligned}\hat{A} &= \big(A_0^{-1} + F_f'(\Omega_1^*)^{-1}F_f\big)^{-1}\\ \hat{a} &= \hat{A}\Big(A_0^{-1}a_0 + F_f'(\Omega_1^*)^{-1}\big(y - Q_1 g_1 - \sum_{j=2}^{J} M_{00}Q_J g_J\big)\Big).\end{aligned}$$

The trick in *Remark 1* will not work on centred functions in step 1 since $Q_{qj}'M_0Q_{qj}$ is not banded. As shown in Jeliazkov and Lee (2010), an application of Sherman-Morrison formulae will ease the computational cost, which is explained in *Remark 2*.

Remark 2: Sampling of Centered Functions. To avoid inverting $\hat{G}_{qj}^{-1}$, the definition of residual maker matrix M_0 is substituted in $\hat{G}_{qj}$. For simplicity, let $\hat{G}_{qj}$ is defined as

$$\hat{G}_{qj} = \Big(\frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\sigma_{i|\{-i\}}^2}Q_{qj}'M_0Q_{qj}\Big)^{-1}$$

After substitution, the equation looks like

$$\hat{G}_{qj} = \Big(\frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\sigma_{i|\{-i\}}^2}Q_{qj}'Q_{qj} - \frac{1}{\sigma_{i|\{-i\}}^2 T}c_{qj}'c_{qj}\Big)^{-1}$$

where $c_{qj} = Q_{qj}'l$. Let $A_{qj} = \frac{1}{\tau_{qj}^2}K_{qj} + \frac{1}{\sigma_{i|\{-i\}}^2}Q_{qj}'Q_{qj}$, $u_{qj} = \frac{c_{qj}}{\sqrt{\sigma_{i|\{-i\}}^2 T}}$ and $\lambda_{qj} = u_{qj}'A_{qj}u_{qj}$. Using the Sherman-Morrison formulae, $\hat{G}_{qj}$ can be written as

$$\hat{G}_{qj} \quad = \quad A_{qj}^{-1} + \frac{A_{qj}^{-1}u_{qj}u_{qj}'A_{qj}^{-1}}{1-\lambda_{qj}}$$

Efficiency gains are achieved as $\hat{g}_{qj}$ can be obtained without inverting A_{qj}. Let $B_{qj} = (A_{qj} + \frac{u_{qj}u_{qj}'}{1-\lambda_{qj}})$, and thus $\hat{G}_{qj} = A_{qj}^{-1}B_{qj}A_{qj}^{-1}$. Following the below-mentioned steps, the sampling of g_{qj} from $N(\hat{g}_{qj},\hat{G}_{qj})$ can be done through $O(T)$ operations instead of $O(T^3)$.

STEP 1. Draw $w_1 \sim N(0,A_{qj})$ and $w_2 \sim N(0,1)$.
STEP 2. Let $w_3 = w_1 + w_2 u_{qj}\sqrt{1-\lambda_{qj}}$ so that $w_3 \sim N(0,B_{qj})$.
STEP 3. Let $w_4 = A_{qj}w_3$ so that $w_4 \sim N(0,\hat{G}_{qj})$.
STEP 4. Let $g_{qj} = \hat{g}_{qj} + w_4$ so that $g_{qj} \sim N(\hat{g}_{qj},\hat{G}_{qj})$.

6. Model Comparison

Since the researcher's take on a specific problem is reflected in how a model is specified and estimated, it is important to know which one accounts better for uncertainty. Especially problems involving empirical analysis, which aims to explain an economic phenomenon and have multiple models under their disposal. Given models M_i and M_k, Bayesian formulation provides a straightforward way to compare models using posterior odds, which involves the prior odds and ratio of marginal likelihoods (Bayes factor). The posterior odds take the following form

$$\frac{p(M_i|y)}{p(M_k|y)} = \frac{p(M_i)\, m(y|M_i)}{p(M_k)\, m(y|M_k)}$$

where the marginal likelihood $m(y|M_i)$ is defined as

$$m(y|M_i) = \int p(y|\theta_i, M_i) p(\theta_i|M_i) d\theta_i$$

where $p(y|M_i, \theta_i)$ is the likelihood for M_i and $p(\theta_i|M_i)$ is the priors on the parameter vector θ_i used in M_i. The additive function framework used in NPVAR-DF model makes θ_i high dimensional, making numerical integration highly costly. Using the application of Bayes theorem, Chib (1995) suggested a more tractable formula to calculate the marginal likelihood, which is given as

$$m(y|M_i) = \frac{p(y|\theta_i^*, M_i) p(\theta_i^*|M_i)}{p(\theta_i^*|y, M_i)}$$

where θ_i^* is the posterior mean. In the case of the NPVAR-DF model, the marginal likelihood will be calculated as

$$m(y) = \frac{p(y|g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*) p(g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*)}{p(g^*, \tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*|y)}$$

given the Gaussian structure of the model, the marginal likelihood can be calculated as

$$m(y) = \frac{p(y|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*) p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*)}{p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*|y)} \tag{20}$$

where densities are marginalized over g. This is possible since conditional on other parameters, the density $p(y|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*)$ is still normal (Chib and Jeliazkov (2006); Jeliazkov (2013)). This helps avoid evaluating the densities at high dimensional g, saving computational costs. The numerator in Eq. 20 is readily available as defined in this paper. To evaluate the posterior density at the posterior mean value, the paper uses the law of probability as

$$\begin{aligned} p(\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, \delta^*|y) = p(\tau^{2*}|y) p(\Omega_1^*|\tau^{2*}, y) p(a^*|\Omega_1^*, \tau^{2*}, y) \dots \\ p(\delta^*|\tau^{2*}, \Omega_1^*, a^*, f^*, \gamma^*, \sigma^{2*}, y) \end{aligned} \tag{21}$$

Chib (1995) provided a method of calculating marginal likelihood under Gibbs sampling. Since the full conditionals of $\tau^2, \Omega_1, f, a, \sigma^2$ and δ are known, the marginal densities in Eq. 21 and g can be estimated using Rao-Blackwellization (Tanner and Wong (1987); Gelfand and Smith (1990)). For given parameters θ_i and θ_k, the conditional density of θ_i marginal of θ_k can be calculated as

$$p(\theta_i|y) = \frac{1}{B}\sum_{b=1}^{B} p(\theta_i|\theta_k^{\{b\}}, y)$$

using B draws of θ_k from the MCMC run. For the estimation of the marginal density of γ, one can refer to Chib and Jeliazkov (2001, 2005).

7. Model Extensions

The estimation technique for the NPVAR-DF model is flexible and readily applicable to various situations. This is possible since the estimation of non-parametric functions and factors can be done conditionally on the adjustments done in other parts of the model. This section illustrates two such model extensions which further NPVAR-DF model applicability in familiar settings and support the claim for the model's modularity.

7.1. **NPVAR-DF with Serially Correlated Errors.**

The model in Eq.7 is extended by considering the serially correlated error term that follows a mean zero stationary $AR(N)$ process.

$$\epsilon_{qt} = \rho_1\epsilon_{q,t-1} + \cdots + \rho_N\epsilon_{q,t-N} + \eta_{qt} \tag{22}$$

where $\eta_{qt} \sim N(0,1)$. The prior density for ρ is defined as $\rho \sim N(\rho_0, P_0)I_{S_\rho}$ where $\rho = (\rho_1, \ldots, \rho_N)'$, I_{S_ρ} is the indicator function which takes the value 1 in case $\rho \in S_\rho$ and S_ρ is the set of ρ vectors which satisfy stationarity. Due to the new structure of the variance-covariance matrix, *Remark 2* cannot be implemented, making computing its inverse a very time-consuming task. Following Chib and Jeliazkov (2006); Jeliazkov (2013), the error terms can be orthogonalized by breaking the variance-covariance matrix into correlated and uncorrelated parts. Then, the correlated part is treated as random effects, and the model is tweaked to take the homoskedastic form. To illustrate, lets decompose $\Omega_q = \Sigma_q - \kappa I$ where Σ_q is a symmetric positive definite matrix and $\kappa > 0$. The paper assumes κ to be equal to the minimum of the set of eigenvalues of Ω_q, which ensures that Ω_q can be decomposed (Chib and Jeliazkov (2006)). Furthermore, let $C_q^{'}C_q = \Sigma_q$ where C_q be the Cholesky decomposition of Σ_q and $\Omega_q = C_q^{'}C_q - \kappa I$. The model in Eq.7 can be rewritten as

$$\begin{aligned}
y_1 &= Q_{11}g_{11} + Q_{12}g_{12} \cdots + Q_{1J}g_{1J} + A_1 f + C_1' u_1 + \xi_1 \\
&\vdots \qquad \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\
y_Q &= Q_{Q1}g_{Q1} + Q_{Q2}g_{Q1} \cdots + Q_{QJ}g_{QJ} + A_Q f + C_Q' u_Q + \xi_Q \\
H_f f &= Z\delta + v
\end{aligned} \tag{23}$$

Stacking over q, the model can be compactly written as

$$\begin{aligned}
y &= Q_1 g_1 + Q_2 g_2 + \cdots + Q_J g_J + Af + C^{'}u + \xi\,, \quad && \xi \sim N(0, \kappa I) \\
H_f f &= Z\delta + v && v \sim N(0, S)
\end{aligned} \tag{24}$$

where $u = \{u_1^{'}, \ldots, u_Q^{'}\}^{'}$ and

$$C = \begin{bmatrix} C_1 & & \\ & \ddots & \\ & & C_Q \end{bmatrix}$$

Given $C^{'}u$, which is computationally inexpensive to calculate, the variance is constant across time since $var(\xi) = \kappa I$. The MCMC algorithm to draw from the posterior distribution will take the following form.

Algorithm 3: MCMC Implementation of the NPVAR-DF Model with Serially-Correlated Errors

STEP 1. For $q = 1, \ldots, Q$

(1) Sample $[u_q|y_q, \theta] \sim N(\hat{u}_q, \hat{U}_q)$ where, $\hat{U}_q = (I + C_q C_q^{'}/\kappa)^{-1}$ and $\hat{u}_q = U_q C_q (y_q - \mu_{q|\{-q\}} - Q_{q1}g_{q1} - \sum_{r\geq 2} M_0 Q_{qr} g_{qr} - A_q f)/\kappa$.

(2) Sample the first function (non-centered) in equation q as $[g_{q1}|y, \theta_{-g_{q1}}] \sim N(\hat{g_{q1}}, \hat{G_{q1}})$ where

$$\begin{aligned} \hat{G}_{q1} &= \Big(\frac{1}{\tau_{q1}^2} K_{q1} + \frac{1}{\kappa} Q_{q1}^{\prime} Q_{q1}\Big)^{-1} \\ \hat{g}_{q1} &= \hat{G}_{q1}\Big(\frac{1}{\tau_{q1}^2} K_{q1} g_{q10} + \frac{1}{\kappa} Q_{q1}^{\prime}\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj} g_{qj} - A_q f \\ &\quad - C_q u_q\big)\Big) \end{aligned}$$

(3) The centered functions are sampled as $[g_{qj}|y, \theta_{-g_{qj}}] \sim N(\hat{g_{qj}}, \hat{G_{qj}})$ for $j = 2, \ldots, J$ where,

$$\begin{aligned} \hat{G}_{qj} &= \Big(\frac{1}{\tau_{qj}^2} K_{qj} + \frac{1}{\kappa} Q_{qj}^{\prime} M_0 Q_{qj}\Big)^{-1} \\ \hat{g}_{qj} &= \hat{G}_{qj}\Big(\frac{1}{\tau_{qj}^2} K_{qj} g_{qj0} + \frac{1}{\kappa} Q_{q1}^{\prime} M_0 \big(y_q - \mu_{q|\{-q\}} - \sum_{r\geq 2, k\neq j} M_0 Q_{qr} g_{qr} - \\ &\quad Q_{q1} g_{q1} - A_q f - C_q u_q\big)\Big). \end{aligned}$$

.

STEP 2. Sample $[\tau_{qj}^2|g_{qj}]$ as in Algorithm 2.

STEP 3. Sample $[a, f|y, \theta_{-\{a,f\}}]$ in one block as

(1) Sample $[a|y, \theta_{-\{a,f\}}]$ marginal of f using the Metropolis Hasting algorithm with tailored proposal $a^* \sim q(\tilde{a}, W)$ and accept a^* with probability

$$\alpha_{MH}(a, a^*) = min\left\{1, \frac{\pi(a^*|\theta_{-\{a,f\}})q(a|\tilde{a}, W)}{\pi(a|\theta_{-\{a,f\}})q(a^*|\tilde{a}, W)}\right\}$$

(2) Sample $[f|y, \theta_{-\{f\}}] \sim N(\hat{f}, \hat{F})$ where

$$\hat{F} = \left(K_f + \frac{A'A}{\kappa}\right)^{-1}$$

$$\hat{f} = \hat{F}\Big(K_f \tilde{f} + \frac{A'}{\kappa}\big(y - Q_1 g_1 - \sum_{j=2}^{J} M_{00} Q_J g_J - C' u\big)\Big)$$

STEP 4. Sample $[\gamma_d|f_k, \sigma_k^2]$ as in Algorithm 2.

STEP 5. Sample $[\sigma_d^2|f_k, \gamma_k]$ as in Algorithm 2.

STEP 6. Sample $[\delta|y, \theta_{-\delta}]$ as in Algorithm 2.

STEP 7. Sample $[\rho|y, \theta_{-\rho} \propto \phi(\rho) N(\hat{\rho}, P) I_{S_\rho}$ where $\hat{\rho} = P(P_0^{-1}\rho_0 + E'e)$, $P = (P_0^{-1} + E^{'}E)^{-1}$, and $\phi(\rho) = |\Omega_\rho|^{-Q/2} exp\{-\frac{1}{2}\sum_{q=1}^{Q} e_{q1}^{'}\Omega_\rho^{-1} e_{q1}\}$. Using Metropolis-Hasting algorithm, the proposal (say ρ^*) is drawn from $N(\hat{\rho}, P)I_{S_\rho}$ (Chib and Greenberg (1994)) and accepted with the probability

$$\alpha_{MH}(\rho, \rho^*) = min\Big\{1, \frac{\phi(\rho^*)}{\phi(\rho)}\Big\}$$

where, $e_{qt} = y_{qt} - g_{q1}(s_{q1t}) + \cdots + g_{qJ}(s_{qJt}) + a_q^{'} f_t$, $e_q = (e_{q,N+1}, \ldots, e_{q,T})'$, $e = (e_1^{'}, \ldots, e_Q^{'})^{'}$ and E denote the $(T-N)Q \times N$ matrix with rows containing N lags of $e_{qt} = (e_{q,t-1}, \ldots, e_{q,t-N})$ and $t \geq N+1$. The initial N values of e_{qt} corresponding to each equation is $e_{q1} = (e_{q1}, \ldots, e_{qN})'$. The matrix Ω_ρ is $N \times N$ stationary covariance matrix constructed exactly as Ω_q.

Now on the construction of Ω_q, let $\epsilon_q = (\epsilon_{q1}, \ldots, \epsilon_{qT})'$ where $\epsilon_q \sim N(0, \Omega_q)$. Given the AR(N) structure of ϵ_{qt} in Eq. 22, Ω_q is a $T \times T$ dimensional Toeplitz matrix. Generally, Ω_q can be determined in the following way. Let $\varphi_n = E(\epsilon_{qt}\epsilon_{q,t-n})$ be the n^{th} autocovariance term. Due to symmetry, $\varphi_n = \varphi_{-n}$. It can be shown that the autocovariance itself follow an AR(N) process as $\varphi_n = \rho_1\varphi_{n-1} + \cdots + \rho_N\varphi_{n-N}$. The first N values $(\varphi_0, \ldots, \varphi_{N-1})$ are given by the first N elements of the first column of the $N^2 \times N^2$ matrix $[I - F \otimes F]^{-1}$ where F, a $N \times N$ matrix, is given as

$$F = \begin{bmatrix} & \rho^{'} & \\ & & \\ I_{N-1} & & 0_{(N-1)\times 1} \end{bmatrix}$$

Thus, Ω_q can be obtained as $\Omega_q[i,k] = \varphi_{i-k}$. To explain it better, if the process considered is AR(1) process, $\Omega_q[i,k] = \rho^{|i-k|}/(1-\rho^2)$.

7.2. **NPVAR-DF with Binary Variable.** Binary, or qualitative, variables have been extensively utilized in Vector Autoregression (VAR) models. A prevalent example of such a variable is a dummy representing business cycle recessions and expansions. Dueker (2005) pioneered the introduction of a Markov Chain Monte Carlo (MCMC) technique for estimating qualitative variables within the VAR framework as dependent variables rather than merely as controls. This development involves incorporating binary variables into the NPVAR-DF model. Denote the binary variable as d_t and the latent variable as y_t^*. Although this section focuses on a single binary variable for simplicity, the methodology can be expanded to include multiple binary variables. The relationship between the binary and latent variables is modelled as follows:

$$d_t = 1 \;\; if \;\; y_t^* > 0$$
$$d_t = 0 \;\; if \;\; y_t^* \leq 0 \tag{25}$$

Considering Eq. 2, the latent variable will enter the model as

$$\begin{aligned}
y_{1t} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{1qpt} + a_1' f_t + \sum_{p=1}^{P} b_{1p} y_{t-p}^* + \epsilon_{1t} \\
&\vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
y_{Qt} &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{Qqpt} + a_Q' f_t + \sum_{p=1}^{P} b_{Qp} y_{t-p}^* + \epsilon_{Qt} \\
y_t^* &= \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{Q+1,qpt} + a_{Q+1}' f_t + \sum_{p=1}^{P} b_{Q+1,p} y_{t-p}^* + \epsilon_{Q+1,t}
\end{aligned} \tag{26}$$

Stacking over variables, Eq. 26 can be written as

$$Y_t \;\; = \;\; \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qpt} + a^* f_t + \sum_{p=1}^{P} b_p y_{t-p}^* + \epsilon_t \tag{27}$$

where,

$$Y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Qt} \\ y_t^* \end{bmatrix}, \; g_{qpt} = \begin{bmatrix} g_{1qpt} \\ \vdots \\ g_{Q+1,qpt} \end{bmatrix}, \; a^* = \begin{bmatrix} a_1' \\ \vdots \\ a_{Q+1}' \end{bmatrix} \;, and \;\; b_p = \begin{bmatrix} b_{1p} \\ \vdots \\ b_{Q+1,p} \end{bmatrix}.$$

The full conditional distribution of y_t^* can be obtained similarly to probit models. Since, the variance of $y^* = (y_1^*, \ldots, y_T^*)$ is not identified, the lower right element of $var(\epsilon_t)$ is kept equal to 1, that is

$$\Omega_1 = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1Q} & \omega_{1,Q+1} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2Q} & \omega_{2,Q+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{Q1} & \omega_{Q2} & \cdots & \omega_{QQ} & \omega_{Q,Q+1} \\ \omega_{Q+1,1} & \omega_{Q+1,2} & \cdots & \omega_{Q+1,Q} & 1 \end{bmatrix}.$$

The latent variable y_t^* follows a truncated normal distribution.

$$d_t = 1 \;\; then \;\; y_t^* \sim TN_{[0,\infty]}$$
$$d_t = 0 \;\; then \;\; y_t^* \sim TN_{[-\infty,0]}. \tag{28}$$

Let Y_{-t} be the full vector time series except for time t. To derive the full conditional density $p(y_t^*|Y_{-t}, y_{1t}, \ldots, y_{Qt}, \theta) = p(y_t^*|\theta^*)$, the density associated with the P lags will be exploited to get a conjugate form. The error terms associated with these P densities are as follows

$$
\begin{aligned}
\epsilon_t &= Y_t - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qpt} - a^* f_t - \sum_{p=1}^{P} b_p y^*_{t-p} \\
\epsilon_{t+1} &= Y_{t+1} - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qp,t+1} - a^* f_{t+1} - \sum_{p=1}^{P} b_p y^*_{(t+1)-p} \\
&\vdots \\
\epsilon_{t+P} &= Y_{t+P} - \sum_{q=1}^{Q}\sum_{p=1}^{P} g_{qp,t+P} - a^* f_{t+P} - \sum_{p=1}^{P} b_p y^*_{(t+P)-p}
\end{aligned} \tag{29}
$$

Let ψ_t be the known part of ϵ_t conditional on $[Y_{-t}, y_{1t}, \ldots, y_{Qt}, \theta]$, then

$$
\begin{aligned}
\epsilon_t &= Y_t - \psi_t \\
\epsilon_{t+1} &= \psi_{t+1} - b_1 y^*_t \\
&\vdots \\
\epsilon_{t+P} &= \psi_{t+P} - b_P y^*_t
\end{aligned} \tag{30}
$$

The conditional density of y^*_t is a function of $(\epsilon_t, \ldots, \epsilon_{t+P})$ and can be written as

$$
\begin{aligned}
p(y^*_t|\theta^*) &= f\Big(exp\{-\frac{1}{2}(\epsilon_t'\Omega_1^{-1}\epsilon_t + \epsilon_{t+1}'\Omega_1^{-1}\epsilon_{t+1} + \cdots + \epsilon_{t+P}'\Omega_1^{-1}\epsilon_{t+P})\}\Big) \\
&= f\Big(exp\{-\frac{1}{2}((Y_t - \psi_t)'\Omega_1^{-1}(Y_t - \psi_t) + (\psi_{t+1} - b_1 y^*_t)'\Omega_1^{-1}(\psi_{t+1} - b_1 y^*_t) \\
&\quad + \cdots + (\psi_{t+P} - b_p y^*_{t+p})'\Omega_1^{-1}(\psi_{t+P} - b_p y^*_{t+P})\}\Big)
\end{aligned}
$$

Specifically, the density of $\epsilon_{Q+1,t}$ will be used in the conditional density instead of ϵ_t and thus $p(y^*_t|\theta^*)$ can be written as

$$
\begin{aligned}
p(y^*_t|\theta^*) &= f\Big(exp\{-\frac{1}{2}((y^*_t - \psi^*_t)'\Omega_{y^*}^{-1}(y^*_t - \psi^*_t) + (\psi_{t+1} - b_1 y^*_t)'\Omega_1^{-1}(\psi_{t+1} - b_1 y^*_t) \\
&\quad + \cdots + (\psi_{t+P} - b_p y^*_{t+p})'\Omega_1^{-1}(\psi_{t+P} - b_p y^*_{t+P})\}\Big)
\end{aligned}
$$

where, $\Omega_{y^*}^{-1} = var(\epsilon_{Q+1,t}|\epsilon_{1,t}, \ldots, \epsilon_{Q,t})$ and $\psi^*_t = E(\epsilon_{Q+1,t}|\epsilon_{1,t}, \ldots, \epsilon_{Q,t})$. Given the normal distribution, obtaining a conditional distribution from joint distribution is straightforward. After collecting all the cross terms, it can be shown that the conditional density for y^*_t will take the following form

$$
[y^*_t|\theta^*] \sim N(C^{-1}D, C^{-1}) \tag{31}
$$

where, $C = (\Omega_{y^*}^{-1} + b_1'\Omega_1^{-1}b_1 + \cdots b_P'\Omega_1^{-1}b_P)$ and $D = -\Omega_{y^*}^{-1}\psi^*_t + b_1'\Omega_1^{-1}\psi_{t+1} + \cdots + b_P'\Omega_1^{-1}\psi_{t+P}$. The latent variable y^*_t is drawn from the truncated normal distribution using Eq. 31 depending upon whether $d_t = 0$ or 1.

The draw of the coefficient b from its conditional distribution can be done by considering the representation in Eq. 7 as

$$
\begin{aligned}
y_1 &= Q_{11}g_{11} + \cdots + Q_{1J}g_{1J} + A_1 f + \tilde{y}^* b_1 + \epsilon_1 \\
&\vdots \qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
y_Q &= Q_{Q1}g_{Q1} \cdots + Q_{QJ}g_{QJ} + A_Q f + \tilde{y}^* b_Q + \epsilon_Q \\
y^* &= Q_{Q+1,1}g_{Q+1,1} \cdots + Q_{Q+1,J}g_{Q+1,J} + A_{Q+1} f + \tilde{y}^* b_{Q+1} + \epsilon_{Q+1}
\end{aligned} \tag{32}
$$

where

$$
\tilde{y}^* = \begin{bmatrix} y_0^* & \cdots & y_{-(P-1)}^* \\ \vdots & \ddots & \vdots \\ y_{T-1}^* & \cdots & y_{T-P}^* \end{bmatrix} \; and \; b_q = \begin{bmatrix} b_{1q} \\ \vdots \\ b_{Pq} \end{bmatrix}
$$

The conditional distribution for b_q will take a similar form as any OLS coefficient. Given the prior as $p(b_q) = N(b_{0q}, B_{0q})$, the conditional density is given as $[b_q|y, \theta_{-\{b_q\}}] \sim N(\hat{b_q}, \hat{B}_q)$ where for $q = 1, \ldots, Q$

$$
\begin{aligned}
\hat{B}_q &= \Big(B_{0q}^{-1} + \frac{1}{\omega_{q|\{-q\}}}\tilde{y}^{*'}\tilde{y}^*\Big)^{-1} \\
\hat{b}_q &= \hat{B}_q^{-1}\Big(B_{0q}^{-1}b_{0q} + \frac{1}{\omega_{q|\{-q\}}}\tilde{y}^{*'}\big(y_q - \mu_{q|\{-q\}} - \sum_{j=2}^{J} M_0 Q_{qj}g_{qj} - Q_{q1}g_{q1} - A_q f\big)\Big)
\end{aligned}
$$

and for $q = Q + 1$

$$
\begin{aligned}
\hat{B}_{Q+1} &= \Big(B_{0,Q+1}^{-1} + \frac{1}{\omega_{Q+1|\{-(Q+1)\}}}\tilde{y}^{*'}\tilde{y}^*\Big)^{-1} \\
\hat{b}_{Q+1} &= \hat{B}_{Q+1}^{-1}\Big(B_{0,Q+1}^{-1}b_{0,Q+1} + \frac{1}{\omega_{Q+1|\{-(Q+1)\}}}\tilde{y}^{*\prime}\big(y^* - \mu_{Q+1|\{-(Q+1)\}} - Q_{Q+1,1}g_{Q+1,1} - \\
&\qquad \sum_{j=2}^{J} M_0 Q_{Q+1,j}g_{Q+1,j} - A_{Q+1} f\big)\Big).
\end{aligned}
$$

8. Application to US Macroeconomic Data

The current application contains U.S. post-war quarterly macroeconomic data from 1948:Q1 to 2005:Q1. The set of variables includes output growth (y_t) measured by log differences of real GDP between two consecutive quarters, average quarterly unemployment rate (u_t), inflation rate (π_t) measured by the percentage change in the Consumer Price Index between consecutive quarters, and interest rates (i_t) measured by the average quarterly secondary market yield on the 3-month Treasury bill. The variables y_t, u_t and i_t are seasonally adjusted. All four variables are summarized in Tab. 1.

NPVAR-DF model is estimated using this data, which helps us examine the behaviour of the dynamic system. The baseline NPVAR-DF(1) model is estimated to achieve parsimony which is given

Table 1. Summary Statistics

Variables	Mean	SD	Min	Max
y_t	0.85	1.00	-2.76	4.02
u_t	5.63	1.52	2.60	10.70
i_t	4.81	2.92	0.79	15.05
π_t	0.92	0.85	-1.24	4.08

as

$$y_t = g_{yy1t}(y_{t-1}) + g_{yu1t}(u_{t-1}) + g_{yi1t}(i_{t-1}) + g_{y\pi1t}(\pi_{t-1}) + a_y f_t + \epsilon_{yt}$$
$$u_t = g_{uy1t}(y_{t-1}) + g_{uu1t}(u_{t-1}) + g_{ui1t}(i_{t-1}) + g_{u\pi1t}(\pi_{t-1}) + a_u f_t + \epsilon_{ut}$$
$$i_t = g_{iy1t}(y_{t-1}) + g_{iu1t}(u_{t-1}) + g_{ii1t}(i_{t-1}) + g_{i\pi1t}(\pi_{t-1}) + a_i f_t + \epsilon_{it}$$
$$\pi_t = g_{\pi y1t}(y_{t-1}) + g_{\pi u1t}(u_{t-1}) + g_{\pi i1t}(i_{t-1}) + g_{\pi\pi1t}(\pi_{t-1}) + a_\pi f_t + \epsilon_{\pi t}$$
$$f_t = \gamma f_{t-1} + v_t\,, \quad \forall \quad t = 1, 2, \ldots, T$$

The model has sixteen non-linear functions and one factor. The Z matrix is not included in the factor equation; thus, its interpretation is heavily tied to the identification restriction. For the identification strategy on the loadings, the model restricts $a_y = 1$. Since a_y is associated with the output growth equation, the factor itself can be interpreted as the business cycle component in the economy, which affects all the macro variables. Dropping Z from the model ensures that the factor has a mean of zero.

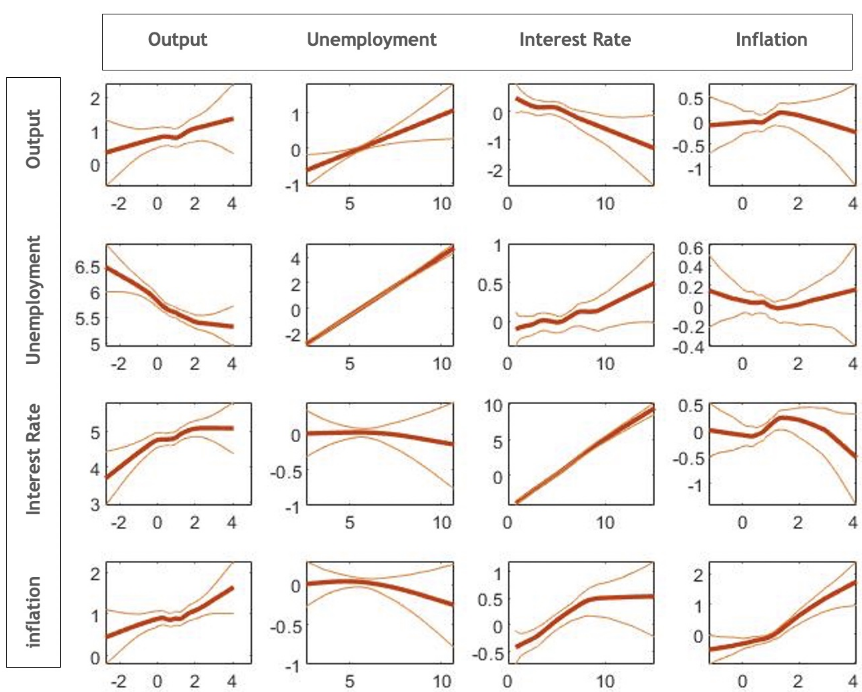

Figure 1. Non parametric functions obtained from estimating NPVAR-DF(1) on US post war data. The Y-axis represents equations and X-axis represents functions.

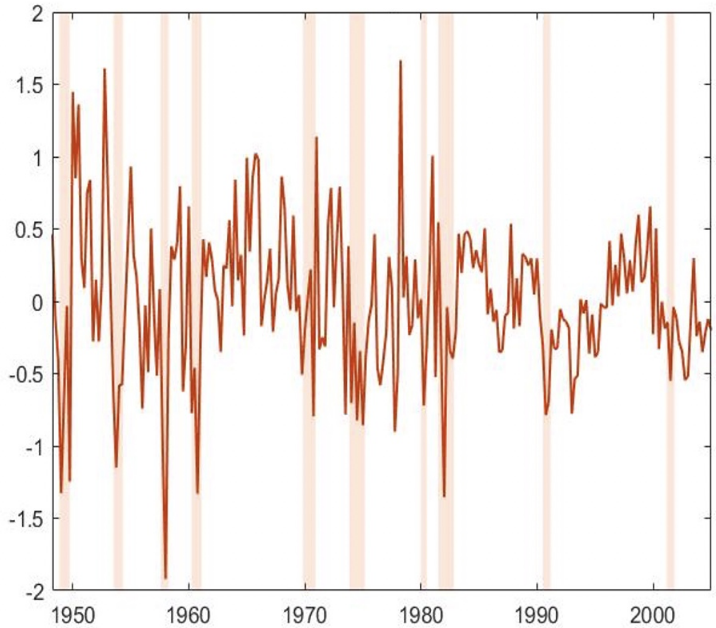

Figure 2. Factor associated with the business cycle component for the US post war data.

Figure 1 presents the estimated functions for the NPVAR-DF(1) model. The bold red line is the mean of the non-parametric functions, whereas the light yellow lines represent 95% credible intervals. Many of the relations, especially the lag effect of unemployment on other macroeconomic variables, can be modelled linearly. To a lesser extent, some of the relations can be approximated as linear. For example, the relation between lag interest rates on the output growth rate, as the loss of information due to linearity restriction, is not significant. However, in many equations, the effects of lagged inflation, lagged interest rates, and lagged output growth rate appear to be nonlinear. These results agree with the literature that has found non-linearity in behaviours of output growth (Dahl and Gonzalez-Rivera (03 2); Dahl and González-Rivera (03 1)) and financial markets variables like interest rates and inflation rate Härdle and Tsybakov (1997); Härdle et al. (1998a)). One advantage of using the second-order Markov process prior for g_{qj} is that the linear relationship is preserved if the true relationship is linear. The prior penalizes deviations from linearity, so the model will spit out non-linear relations only if enough evidence is there to support it.

The credible intervals are shaped as hourglass since the identification restrictions make the intervals narrower at the point where non-parametric functions are centred. Other identification techniques will have a different effect on the shape of credible intervals. For example, forcing all but one non-parametric function to start from 0 would have led to funnel-shaped credible intervals.

The dynamic factor estimated in this application is shown in Fig 2. The factor is superimposed over the shaded periods, representing officially announced recessions by the Fed in the US over the period considered. The factor is able to capture the business cycle element of the economy as it coincides with the recessions. There are specific periods where the factor is off even though the economy is not going through the recession (around 1978, 1993 and 2002). One of the explanations is that there are cases of decline in GDP growth rate (still positive), in these periods, especially from 1978 Q3 - 1979 Q1. The factor can be used in future research to represent the business cycle component in a macro model.

9. Conclusion

This paper has presented the specification, identification, and estimation of non-parametric VAR models with dynamic factors. The paper proposes an efficient MCMC sampling algorithm to estimate the model efficiently. The paper expands on model comparison and extensions on the model. The model is able to accommodate the error terms following AR(1) process, but the idea can be generalized to any form of heteroskedasticity or autocorrelation, which can be taken up in future endeavours. The model's extension with qualitative variables expands its applicability to datasets with binary variables. The application considered U.S. post-war data on GDP growth, unemployment, interest rates, and inflation. The model was able to recover non-linear relationships concerning the financial variables and the factor was able to capture the business cycle component in the economy. Due to the modality of the NPVAR-DF model, it can be extended to consider non-normal error terms and applied to other economies.

References

Amir-Ahmadi, P. and Uhlig, H. (2009). Measuring the dynamic effects of monetary policy shocks: A bayesian favar approach with sign restriction.

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of favar models. *Journal of Business & Economic Statistics*, 34(4):620–641.

Balke, N. S. (2000). Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, 82(2):344–349.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Canova, F. (1993). Modelling and forecasting exchange rates with a bayesian time-varying coefficient model. *Journal of Economic Dynamics and Control*, 17(1-2):233–261.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association*, 90(432):1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75(1):79–97.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183–206.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American statistical association*, 96(453):270–281.

Chib, S. and Jeliazkov, I. (2005). Accept–reject metropolis–hastings sampling and marginal likelihood estimation. *Statistica Neerlandica*, 59(1):30–44.

Chib, S. and Jeliazkov, I. (2006). Inference in semiparametric dynamic models for binary longitudinal data. *Journal of the American Statistical Association*, 101(474):685–700.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics*, 8(2):262–302.

Dahl, C. M. and González-Rivera, G. (2003-1). Testing for neglected nonlinearity in regression models based on the theory of random fields. *Journal of Econometrics*, 114(1):141–164.

Dahl, C. M. and Gonzalez-Rivera, G. (2003-2). Identifying nonlinear components by random fields in the us gnp growth. implications for the shape of the business cycle. *Studies in Nonlinear Dynamics & Econometrics*, 7(1).

Dueker, M. (2005). Dynamic forecasts of qualitative variables: a qual var model of us recessions. *Journal of Business & Economic Statistics*, 23(1):96–104.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 68(1):189–196.

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

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of policy Modeling*, 14(4):517–533.

Härdle, W., Mammen, E., and Müller, M. (1998a). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, 93(444):1461–1474.

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*, 81(1):223–242.

Härdle, W., Tsybakov, A., and Yang, L. (1998b). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, 68(2):221–245.

Hastie, T. and Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, pages 1005–1016.

Huang, B.-N., Hwang, M.-J., and Peng, H.-P. (2005). The asymmetry of the impact of oil price shocks on economic activities: an application of the multivariate threshold model. *Energy Economics*, 27(3):455–476.

Jeliazkov, I. (2013). Nonparametric vector autoregressions: Specification, estimation, and inference. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Jeliazkov, I. and Lee, E. H. (2010). Mcmc perspectives on simulated likelihood estimation. In *Maximum simulated likelihood methods and applications*. Emerald Group Publishing Limited.

Kalli, M. and Griffin, J. E. (2018). Bayesian nonparametric vector autoregressive models. *Journal of econometrics*, 203(2):267–282.

Kim, C.-J., Morley, J., and Piger, J. (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics*, 20(2):291–309.

Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80(2):188–201.

Lindley, D. (1971). Bayesian statistics: A review (siam).

Poirier, D. J. (1998). Revising beliefs in nonidentified models. *Econometric theory*, 14(4):483–509.

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

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3):185–207.

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168.

Sims, C. A. et al. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, 10(Win):2–16.

Sims, C. A. and Zha, T. (2006). Were there regime switches in us monetary policy? *American Economic Review*, 96(1):54–81.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbooks Online*.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.

Tong, H. (1978). On a threshold model.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3):245–268.

Van Robays, I. (2016). Macroeconomic uncertainty and oil price volatility. *Oxford Bulletin of Economics and Statistics*, 78(5):671–693.

Vieu, P. (1995). Order choice in nonlinear autoregressive models. *Statistics: A Journal of Theoretical and Applied Statistics*, 26(4):307–328.

Vigfusson, R. (1997). Switching between chartists and fundamentalists: a markov regime-switching approach. *International Journal of Finance & Economics*, 2(4):291–305.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):364–372.