NONPARAMETRIC VECTOR AUTOREGRESSIONS: SPECIFICATION, ESTIMATION, AND INFERENCE

IVAN JELIAZKOV*
University of California, Irvine

July, 2013

Abstract

For over three decades, vector autoregressions have played a central role in empirical macroeconomics. These models are general, can capture sophisticated dynamic behavior, and can be extended to include features such as structural instability, time-varying parameters, dynamic factors, threshold-crossing behavior, and discrete outcomes. Building upon growing evidence that the assumption of linearity may be undesirable in modeling certain macroeconomic relationships, this paper seeks to add to recent advances in VAR modeling by proposing a nonparametric dynamic model for multivariate time series. In this model, the problems of modeling and estimation are approached from a hierarchical Bayesian perspective. The article considers the issues of identification, estimation, and model comparison, enabling nonparametric VAR models to be fit efficiently by Markov chain Monte Carlo algorithms and compared to parametric and semiparametric alternatives by marginal likelihoods and Bayes factors. Among other benefits, the methodology allows for a more careful study of structural instability while guarding against the possibility of unaccounted nonlinearity in otherwise stable economic relationships. Extensions of the proposed nonparametric model to settings with heteroskedasticity and other important modeling features are also considered. The techniques are employed to study the post-war US economy, confirming the presence of distinct volatility regimes and supporting the contention that certain nonlinear relationships in the data can remain undetected by standard models.

Keywords: Additive model; Vector autoregressive (VAR) model; Bayesian model comparison; Markov chain Monte Carlo.
JEL Codes: C11, C14, C15, C32, C52, E31, E32, E37, E43, E47.

1 Introduction and Motivation

Following the seminal work of Sims (1980), the vector autoregressive (VAR) model has played a central role in empirical macroeconomics. The basic model postulates that a q-dimensional vector of time-series

*Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100. E-mail: ivan@uci.edu. I am grateful to Tom Fomby, Lutz Killian, Anthony Murphy, two anonymous referees, and my colleagues Dale Poirier, David Brownstone, Fabio Milani, and especially Angela Vossmeyer, for their careful comments on earlier drafts.

variables $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{qt})'$ depends on its past realizations through the specification

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{c}$ is a q-vector of intercepts, $\{\boldsymbol{B}_j\}_{j=1}^{p}$ are $q \times q$ matrices of parameters, $\{\boldsymbol{y}_{t-j}\}_{j=1}^{p}$ are lags of $\boldsymbol{y}_t$, and $\boldsymbol{\varepsilon}_t$ is an error term with mean zero and $q \times q$ covariance matrix $\boldsymbol{\Omega}$. VAR models are general and can capture sophisticated dynamic behavior, even when the lag length p is relatively small. Moreover, VAR models are very versatile – in the past few decades they have been adapted to incorporate structural instability, regime switching, time-varying parameters, dynamic factors, threshold-crossing behavior, and discrete data, among others. Consequently, VAR methodology has been an important instrument in policy analysis, forecasting, and academic discourse (for a recent review, see Koop and Korobilis, 2009).

While many extensions of the model in (1) are possible, it has been common practice to maintain a parametric, typically linear, functional form for the conditional mean of y_t given its lags. Important extensions of the basic setup are afforded by models in which the parameters are allowed to change over time as in regime switching, changepoint, time-varying parameter, and threshold models.[1] For instance, following Hamilton (1989), much work has been done on estimating models subject to regime shifts in the mean, variance, or dynamics (e.g., Hansen, 1992; Chib, 1996; Chauvet, 1998; Kim and Nelson, 1999; Kim et al., 2005; Sims and Zha, 2006). Threshold regressions have been considered in Beaudry and Koop (1993), Potter (1995), and Pesaran and Potter (1997), while time-varying parameter applications have been examined in Canova (1993), Stock and Watson (1996), Cogley and Sargent (2001), Primiceri (2005), and Chan and Jeliazkov (2009), among others. Much more rare in time series analysis has been the application of nonparametric methods in the modeling and estimation of the conditional mean of a time series process, and this represents the basic econometric problem motivating this work.

The discussion in this paper is primarily concerned with methods for allowing considerable flexibility in estimating the dependence on lags in VAR models, while maintaining simplicity, computational tractability, and accommodating other modeling features that may be present in the model. The VAR paradigm

[1] In these cases, the modeling involves an additional state variable that can be latent or observed; conditionally on the state, the models are linear, whereas marginalization over the state yields piecewise linear regressions or mixtures of linear regressions.

came to prominence as a methodological framework involving only minimal restrictions. Consequently, continuing in this tradition, this paper discusses ways of estimating multivariate dynamic systems without assuming *a priori* knowledge of the functional form. Even though the nonparametric literature is vast and diverse, applications to time series have been limited despite their potential appeal and importance. In an application to exchange rates, Härdle et al. (1998) use local polynomial methods based on kernel-weighted least squares, to estimate a nonparametric bivariate dynamic system in which both the conditional mean and variance are unknown functions of the past. The methodology in that paper relates to single-equation techniques and exchange rate applications considered in Härdle and Tsybakov (1997) and Yang et al. (1999). Univariate nonparametric regressions have been used in Dahl and Gonzalez-Rivera (2003a) to study the evolution of U.S. GNP growth using the method of Hamilton (2001), who employed the techniques to address nonlinearity in the inflation-unemployment trade-off in an example involving the Phillips Curve. Dahl and Gonzalez-Rivera (2003b) apply the methodology to study the evolution of industrial production for a subsample of OECD countries. Their results support the contention that much nonlinearity is neglected if standard linear models are applied in these settings.

These and other papers have provided a growing body of evidence that allowing for various kinds of nonlinearity in empirical macroeconomics can be very valuable in uncovering important features of time-series relationships. Building upon these advances, this paper seeks to add to the literature by examining a nonparametric dynamic model for multivariate time series. Specifically, this paper considers a dynamic system of q regression equations for data $\{\boldsymbol{y}_t\}_{t=1}^{T}$, where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{qt})'$, in which the ith equation ($i = 1, \ldots, q$) is modeled through the additive form (Hastie and Tibshirani, 1990)

$$y_{it} = \sum_{j=1}^{q} \sum_{k=1}^{p} g_{ijk}(y_{j,t-k}) + \varepsilon_{it}, \quad t = 1, \ldots, T, \tag{2}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{qt})' \sim N(\mathbf{0}, \boldsymbol{\Omega})$ and the unknown functions $\{g_{ijk}(y_{j,t-k})\}$ will be modeled and estimated nonparametrically. For this reason, in the remainder of this paper, the specification in (2) will be referred to as a nonparametric VAR (or NPVAR) model. The model in (2) provides a natural extension of the traditional linear VAR model in (1) – relative to its parametric counterpart, the NPVAR model maintains

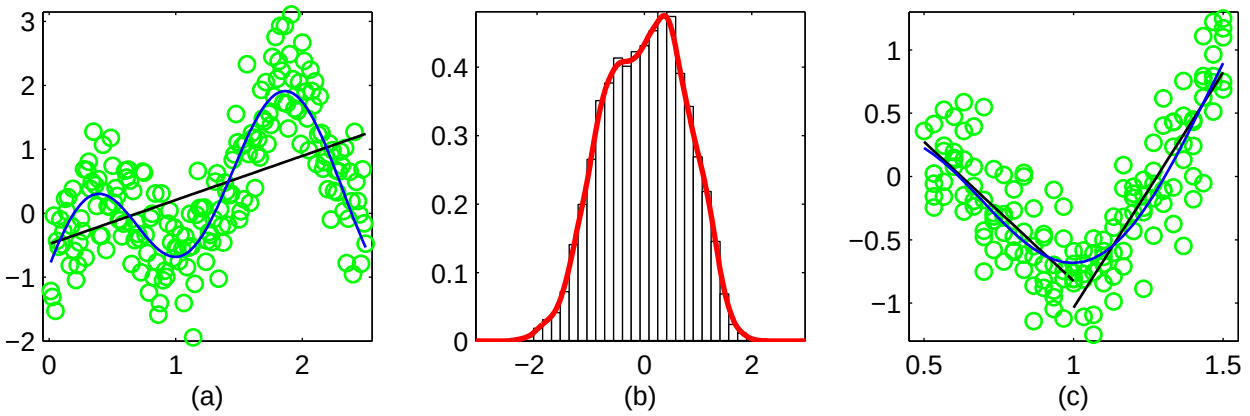
additivity but does not require that the estimated regression relationships lie in a particular class of functions. Functional flexibility is desirable because nonlinearity is common in both economic theory and practice. Moreover, nonparametric additive modeling has desirable practical and theoretical properties and can serve as a useful exploratory tool that is easily inserted in more complex models. Even though the estimation of unknown functions is a complex high-dimensional problem, the additive framework is well suited for dealing with the "curse of dimensionality" because the argument of each function is a single variable.

The specification of the NPVAR model will be approached from a hierarchical Bayesian perspective with special emphasis on the issues of identification, estimation, and model comparison, enabling NPVAR models to be fit efficiently by Markov chain Monte Carlo (MCMC) algorithms and compared to nested and non-nested parametric and semiparametric alternatives by marginal likelihoods and Bayes factors. The methodology is useful in its own right as an exploratory and modeling tool, but is also appealing because it enables a more careful study of other structural features while guarding against the possibility of unaccounted nonlinearity. Doing so is important for theoretical and practical reasons, and because the consequences of ignored nonlinearity can be severe.

The types of misspecification that arise from assuming an inappropriate functional form can be illustrated by considering two simple motivating examples. Imagine that data are generated from the model $y_t = g(x_t) + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$, and $g(\cdot)$ is the nonlinear function in panel (a) of Figure 1. If estimation is by linear regression (the resulting regression line is also shown in the panel (a) of the Figure), it is easy to see that the regression residuals will be heteroskedastic, owing to the neglected nonlinearity in $g(\cdot)$. If the covariate x_t is a lag of y_t, the misspecification can also lead to erroneous findings of serial correlation in the errors. Furthermore, even though the original errors used to generate the data were Gaussian, in a linear regression they will appear non-Gaussian (see panel (b) of Figure 1). Due to the omitted nonlinearity, the error distribution will be a location mixture of normals, and consequently one would conclude that the Gaussian assumption is inadequate when the real culprit is neglected nonlinearity. Note that these problems will not be resolved by using estimators that are robust to distributional misspecification.

For our second example, consider panel (c) of Figure 1. In this case, imagine that the researcher is

Figure 1: Ignored nonlinearity can lead to erroneous conclusions about the presence of heteroskedasticity or autocorrelation, the adequacy of the distributional assumptions, the structural stability of the regression, or lead to conclusions that more profligate models are required.



aware that the data generating process involves a nonlinear mean, but chooses to restrict attention to the class of piecewise linear polynomials. Although heteroskedasticity, autocorrelation, and non-Gaussianity may not be significant (or even discernable) problems if a bilinear model is fit to the particular data in panel (c) of the Figure, one can erroneously conclude that there is evidence of structural instability. For instance, regime switching, changepoint, threshold, and time-varying parameter models may appreciably improve the fit relative to a linear model, although one should bear in mind that these would be spurious findings of instability or structural breaks since the underlying data generating process is a stable, although nonlinear, function of the covariates that is not properly accommodated in the regressions. Such spurious instability, unfortunately, is not the only pitfall that can be induced by this type of misspecification. The poor fit of low order linear dynamic systems may also lead researchers to explore more profligate models involving more lags. This would lead to loss of parsimony as additional lag components are incorporated but simply act as atheoretical fitting parameters in the model.

These examples provide strong motivation for studying the NPVAR model because the problems they identify can not be addressed satisfactorily without directly addressing the flexibility of the functional form. This, of course, is not to say that features such as heteroskedasticity, autocorrelation, non-Gaussianity, or structural instability can not be present in nonlinear models. On the contrary, they can be important

integral parts of the NPVAR model, and many such extensions will be considered in Section 5 and the application in Section 6. However, the examples do suggest that before jumping to conclusions about the presence of any of the aforementioned features, one must ensure that they are not spuriously induced by functional form misspecifcation. To enable this task to be carried out, this paper provides methodology for the specification, estimation, and comparison of nonparametric models, which can be useful in this pursuit, as demonstrated in a study of U.S. macroeconomic data. The application reveals that the NPVAR model supports the existence of distinct volatility regimes in the data, and provides evidence that means remain stable but exhibit interesting nonlinearities.

The remainder of this article deals with the hierarchical structure of NPVAR models and their implementation in practice. Specifically, Section 2 presents the specification of the NPVAR model together with a computationally convenient identification restriction on the unknown additive functions. Section 3 presents an efficient fitting algorithm based on MCMC simulation techniques, which subsumes frequentist estimation by backfitting as a special case. Section 4 addresses the problem of model comparison and model averaging by discussing the computation of marginal likelihoods and Bayes factors. Section 5 outlines extensions to settings with heteroskedasticity, Student-t errors, structural instability, heteroskedasticity or stochastic volatility, dynamic factors, and discrete outcomes, and provides references to the relevant literature. Section 6 considers the application of the NPVAR model to data for the post-war US economy, whereas Section 7 offers concluding remarks.

2 Hierarchical Model Specification

The NPVAR model will be specified through the following distributional hierarchy. The likelihood function $f(\boldsymbol{y}|\{g_{ijk}(\cdot)\}, \boldsymbol{\Omega})$ obtained from the additive model in (2) will be augmented with a prior distribution (or model) for the each unknown function $\pi(g_{ijk}(\cdot)|\tau^2_{ijk})$ that will, in turn, depend on a hyperparameter τ^2_{ijk}. The hierarchy will be completed by the priors on τ^2_{ijk} and $\boldsymbol{\Omega}$, denoted by $\pi(\tau^2_{ijk})$ and $\pi(\boldsymbol{\Omega})$, respectively. The prior $\pi(g_{ijk}(\cdot)|\tau^2_{ijk})$ is often referred to as a "smoothness prior" because it aims at penalizing rough functions but does not absolutely rule out any values that the function can take. The prior is very similar to

the roughness penalty in frequentist penalized likelihood estimation. The parameters τ_{ijk}^2 are often called smoothness parameters because they control the degree of smoothness of $g_{ijk}(\cdot)$ in $\pi(g_{ijk}(\cdot)|\tau_{ijk}^2)$. Details will be provided in the remainder of this section, but it is important to note that the methodology leans on a vast literature in Bayesian nonparametric estimation in a variety of areas with continuous, discrete, and censored responses, including cross-sectional settings (Besag et al., 1995; Wood and Kohn, 1998; Hastie and Tibshirani, 2000; Fahrmeir and Lang, 2001; Wood et al., 2002; Koop and Poirier, 2004), multiple equation systems (Smith and Kohn, 2000; Holmes et al., 2002; Koop et al., 2005), panel data (Chib and Jeliazkov, 2006), sample selection models (Chib and Greenberg, 2007; Chib et al., 2009), and time series applications (Hamilton, 2001). Extensions to Bayesian models with free-knot splines have been pursued in Denison et al. (1998) and DiMatteo et al. (2001), while Bayesian estimation techniques for multivariate functions have been provided in Shively et al. (1999) and Wood et al. (2002). Useful reviews and introduction to many aspects of nonparametric modeling can be found in Hastie and Tibshirani (1990), Denison et al. (2002), Koop (2003), Ruppert et al. (2003), Wasserman (2006), and Ahamada and Flachaire (2010).[2] Nonparametric functional modeling has appealing frequentist and Bayesian properties, and many of its advantages have been illustrated in the aforementioned works.

Some simplification in the notation can be obtained by denoting the $r = qp$ lagged variables on the right hand side of the ith equation ($i = 1, \ldots, q$) in (2) by $\{s_{ijt}\}_{j=1}^{r}$ and writing that equation as

$$y_{it} = g_{i1}(s_{i1t}) + \ldots + g_{ir}(s_{irt}) + \varepsilon_{it}, \quad (t = 1, \ldots, T). \tag{3}$$

Then, to motivate the hierarchical model for the functions, it is useful to stack the observations and write the model in matrix notation. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, and for each of the $j = 1, \ldots, r$ functions in (3), let the T observations in the covariate vectors $\boldsymbol{s}_{ij} = (s_{ij1}, \ldots, s_{ijT})'$ determine the corresponding $m_j \times 1$ *design point vectors* $\boldsymbol{v}_{ij} = \left(v_{ij1}, \ldots, v_{ijm_j}\right)'$ with entries equal the *unique ordered* values of $\boldsymbol{s}_{ij}$, that is $v_{ij1} < \ldots < v_{ijm_j}$. Let the corresponding function evaluation vectors be denoted by $\boldsymbol{g}_{ij} = \left(g_{ij}\left(v_{ij1}\right), \ldots, g_{ij}(v_{ijm_j})\right)'$. Then, stacking over time, the ith equation of the system can be written

[2]Interested readers are referred to these books for further details on a rich variety of nonparametric modeling approaches such as truncated polynomials, radial basis functions, neural networks, regression trees, wavelets, kernel smoothing, locally weighted polynomials, B-splines, etc., many of which are beyond the scope of this paper.

in matrix notation as

$$\boldsymbol{y}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{Q}_{i2}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{Q}_{ir}\boldsymbol{g}_{ir} + \boldsymbol{\varepsilon}_i, \tag{4}$$

where $\boldsymbol{Q}_{ij}$ are $T \times m_{ij}$ *incidence matrices* with entries $\boldsymbol{Q}_{ij}(h,k) = 1$ if $s_{ijh} = v_{ijk}$ and 0 otherwise, which establishes the correspondence between $\boldsymbol{s}_{ij}$ and $\boldsymbol{v}_{ij}$. Note that because there may be repeating values in $\boldsymbol{s}_{ij}$, we have that $m_j \leq T$ for $j = 1, \ldots, r$. Since all rows of $\boldsymbol{Q}_j$ contain a single 1, row t of the product $\boldsymbol{Q}_{ij}\boldsymbol{g}_{ij}$ is given by $g_{ij}(s_{ijt})$.

The idea behind nonparametric modeling is to view the function evaluations in each $\boldsymbol{g}_{ij}$ as the realization of a stochastic process which controls the degree of local variation between neighboring elements. Despite differing theoretical foundations and assumptions, the vectors of function evaluations $\boldsymbol{g}_{ij}$ can eventually be written, in a wide range of nonparametric modeling approaches, as random fields of the form

$$\boldsymbol{g}_{ij}|\tau_{ij}^2 \sim N\left(\boldsymbol{g}_{ij0}, \tau_{ij}^2\boldsymbol{K}_{ij}^{-1}\right), \quad j = 1, \ldots, r, \tag{5}$$

where τ_{ij}^2 is a smoothness parameter and $\boldsymbol{K}_{ij}$ is a matrix whose structure will be discussed shortly. From a Bayesian perspective, equation (5) can be viewed as a smoothness prior for $\boldsymbol{g}_{ij}$, where as from a frequentist perspective, it is often viewed as a roughness penalty term in penalized likelihood estimation (Wahba, 1978). In either case, the goal of the modeling is to introduce a penalty to local variation between successive elements of $\boldsymbol{g}_{ij}$, without absolutely ruling out any possible value that the elements of $\boldsymbol{g}_{ij}$ can take.

The focus in this paper is on models involving banded precision matrices $\boldsymbol{K}_{ij}$, i.e., matrices which have non-zero elements only in small bands around the main diagonal. Matrix bandedness is a feature that significantly reduces the computational costs and makes the analysis of high-dimensional problems feasible and inexpensive. This paper will examine the construction of $\boldsymbol{g}_{ij0}$ and $\boldsymbol{K}_{ij}$ in (5) for a class of smoothness priors, which are conceptually simple and easily adaptable, can approximate unknown functions arbitrarily well, and have been widely used (see, for example, Poirier, 1973; Shiller, 1984; Besag et al., 1995; Fahrmeir and Lang, 2001; Koop and Poirier, 2004; Koop et al., 2005; Chib and Jeliazkov, 2006; Chib et al., 2009). The roots of this method can be traced back to Whittaker (1923), and its relationship with state space models has been discussed in Chan and Jeliazkov (2009). It should be noted that despite the

focus on a specific smoothness prior, the estimation methodology described in this paper is generic and can be applied with various modeling approaches for the unknown functions, such as splines (Poirier, 1973; Shiller, 1984), B-splines (Silverman, 1985), wavelets (Denison et al., 2002), or the approach of Hamilton (2001). Although bandedness of the precision matrix is a useful characteristic of many of the preceding nonparametric approaches, it is not a feature of other popular modeling methods, e.g. regression splines or integrated Wiener process priors, which may be more computationally intensive in high-dimensional cases.

Because the modeling follows identical steps for each of the functions, for the time being we can simplify notation by suppressing the ij subscripts that denote the equation and function numbers. With this convention, a Markov process prior views the elements of $\boldsymbol{g} = (g(v_1), \ldots, g(v_m))' \equiv (g_1, \ldots, g_m)'$ as a stochastic process observed at the unique and ordered values in $\boldsymbol{v}$. Specifically, letting $h_\ell \equiv v_\ell - v_{\ell-1}$, a first-order Markov process prior can be defined as

$$g_\ell = g_{\ell-1} + u_\ell, \tag{6}$$

while a second-order Markov process prior is given by

$$g_\ell = \left(1 + \frac{h_\ell}{h_{\ell-1}}\right) g_{\ell-1} - \frac{h_\ell}{h_{\ell-1}} g_{\ell-2} + u_\ell, \tag{7}$$

where $u_\ell \sim N(0, \tau^2 h_\ell)$ and τ^2 is a smoothness parameter, such that small values of τ^2 produce smoother functions, while larger values allow the function to be more flexible and interpolate the data more closely. The weights h_ℓ adjust the variance to account for possibly irregular spacing between consecutive points in each design vector; the one given here implies that the variance grows linearly with the distance h_ℓ, although other weights are also possible. A distribution for the initial states of the stochastic process is necessary in order to complete the specification of the smoothness prior. For example, for the first-order prior, the initial state can be modeled as

$$g_1 \sim N\left(g_{10}, \tau^2 G_{10}\right), \tag{8}$$

whereas in the second-order case, we have

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} | \tau^2 \sim N\left(\begin{pmatrix} g_{10} \\ g_{20} \end{pmatrix}, \tau^2 \boldsymbol{G}_0\right), \tag{9}$$

where $\boldsymbol{G}_0$ is a 2×2 symmetric positive definite matrix. The prior on the initial conditions in (8) and (9) is very important because it induces a *proper* prior on the remaining observations (see Chib and Jeliazkov, 2006). Specifically, equation (6), starting with the initial condition in (8), implies a penalty on abrupt jumps between successive function evaluations, whereas (7), starting with (9), induces a more general prior on linear functions of $\boldsymbol{v}_j$ that is conceptually similar to the usual priors placed on the intercept and slope parameters in linear regression. This can be seen more precisely by iterating (7) in expectation (to eliminate u_ℓ which is the source of the nonlinearity), starting with initial states in (9).

The interpretability of the directed Markovian structure of the priors specified by (6)–(9) is a convenient aspect of this approach, however, it also leads to an equivalent undirected representation that is used in deriving the random field version of the smoothness prior in (5). This can be shown by recongnizing that upon defining

$$\boldsymbol{H} = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} G_{10} & & & \\ & h_2 & & \\ & & \ddots & \\ & & & h_m \end{pmatrix},$$

for the first-order case in equations (6) and (8), and similarly letting

$$\boldsymbol{H} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ \frac{h_3}{h_2} & -\left(1+\frac{h_3}{h_2}\right) & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{h_m}{h_{m-1}} & -\left(1+\frac{h_m}{h_{m-1}}\right) & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{G}_0 & & & \\ & h_3 & & \\ & & \ddots & \\ & & & h_m \end{pmatrix},$$

for the second-order Markov process in (7) and (9), one can write $\boldsymbol{Hg} = \boldsymbol{u}$, where $\boldsymbol{u} \sim N(\boldsymbol{u}_0, \boldsymbol{\Sigma})$ is used to denote the errors in the Markov process with $\boldsymbol{u}_0 = (g_{10}, 0, \ldots, 0)'$ and $\boldsymbol{u}_0 = (g_{10}, g_{10}, 0, \ldots, 0)'$ in the first- and second-order cases, respectively. A simple change of variables technique leads to the distribution $\boldsymbol{g}|\tau^2 \sim N\left(\boldsymbol{g}_0, \tau^2 \boldsymbol{K}^{-1}\right)$, where the penalty matrix $\boldsymbol{K}$ is given by $\boldsymbol{K} = \boldsymbol{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{H}$ and $\boldsymbol{g}_0 = \boldsymbol{H}^{-1}\boldsymbol{u}_0$. This derivation leads to the distributions presented in (5), where the indices i and j are explicitly present. Note that $\boldsymbol{g}_0$ can alternatively be derived by taking recursive expectations of either (6) or (7) starting with the mean in (8) or (9), respectively.

Two key features of the class of priors are that (i) they are proper, which allows for formal Bayesian

model selection, and (ii) the $m \times m$ penalty matrices $\boldsymbol{K}$ are banded, which is of considerable convenience, as manipulations involving such matrices take $O(m)$ operations, rather than the usual $O(m^3)$ operations for inversions and determinant computations, or $O(m^2)$ operations for multiplication by a vector. Given that m may potentially be as large as the total number of observations T in the sample, this has important ramifications for the numerical efficiency.

Since the priors on $\{\boldsymbol{g}_{ij}\}$ are defined conditionally on the hyperparameters $\{\tau_{ij}^2\}$, the hierarchical structure of the model is completed by specifying the prior distributions $\tau_{ij}^2 \sim IG\left(\nu_{ij0}/2, \delta_{ij0}/2\right)$. Similarly, the prior distribution on the covariance matrix $\boldsymbol{\Omega}$ is taken as $\boldsymbol{\Omega}^{-1} \sim W\left(r_0, R_0\right)$. In setting these priors it is generally very helpful to consider their mapping to the mean and variance of the inverse gamma and Wishart distributions (see Gelman et al., 2003, App. A), as the choice of these parameters plays a role in determining the trade-off between smoothness and goodness of fit. An example of how different settings of the prior parameters can lead to over- or under-smoothing is presented in Chib and Jeliazkov (2006).

Before we can focus on estimating the model, we must address the likelihood identification problem that emerges due to the additive structure in (3). Because the likelihood will remain unchanged if we simultaneously let $g_j^*(\cdot) = g_j(\cdot) + \alpha$ and $g_k^*(\cdot) = g_k(\cdot) - \alpha$ for $k \neq j$, it is obvious that neither an intercept, nor the level of the individual functions is likelihood identified. This can also be seen by recognizing that all rows of every incidence matrix $\boldsymbol{Q}_j$ in (4) sum to 1, leading to perfect multicollinearity because model (4) can be thought of as a saturated dummy variable model. Therefore, the functions must be appropriately "anchored" in order to achieve likelihood identification.

It is well known that Bayesian models with proper priors do not suffer from identification problems even when the likelihood is not identified (Lindley, 1971; Poirier, 1998). However, because the nonparametric components of additive models are correlated by construction (since they enter the mean function additively), likelihood identification is essential for providing a model with well-behaved conditional posterior distributions that will produce quickly-mixing MCMC algorithms for efficient posterior sampling. To achieve likelihood identification, I remove free constants in the likelihood by employing the identification restrictions proposed in Jeliazkov (2011). The approach formally identifies the model by centering the func-

tions in the likelihood and integrating out—instead of holding fixed—any unidentified quantities that enter the specification.[3] Such quantities are marginalized out with respect to a proper prior that is of no relevance in the likelihood and does not affect marginal likelihood estimation, thus relating this approach to the idea of marginal data augmentation discussed in Meng and van Dyk (1999), van Dyk and Meng (2001), and Imai and van Dyk (2005).

One approach to identification is to remove the free constants in the likelihood by restricting $r-1$ of the functions $\{\boldsymbol{g}_{ij}\}$ in each equation to start at zero (e.g., Shively et al., 1999; Koop et al., 2005; Chib et al., 2009). While this approach is quite natural as it corresponds to creating a baseline category in dummy variable models, in the context of nonparametric regression it tends to produce funnel-shaped error bands for the function estimates due to the identification restriction. Consequently, the information content in the data can be confounded with the repercussions of the identification restriction, so that narrower bands need not correspond to regions with more data or better identification of the function.

Another possibility for anchoring the functions is to consider the following version of (4)

$$\boldsymbol{y}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{Q}_{i2}\boldsymbol{M}_{02}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{Q}_{ir}\boldsymbol{M}_{0r}\boldsymbol{g}_{ir} + \boldsymbol{\varepsilon}_i, \tag{10}$$

where

$$\boldsymbol{M}_{0j} = \left(\boldsymbol{I}_{m_j} - \frac{\mathbf{1}_{m_j}\mathbf{1}'_{m_j}}{m_j}\right), \quad j = 1, \ldots, r,$$

are $m_j \times m_j$ symmetric and idempotent mean-differencing matrices (Hastie and Tibshirani, 1990; Lin and Zhang, 1999). Unfortunately, this identification scheme does not lend itself to computationally efficient posterior simulation and, as pointed out by Gelfand (2000), it has been applied in ways that do not correspond to well-defined Bayesian models, with centering typically introduced "on the fly" merely as a step in the fitting algorithm.

For these reasons, this paper employs the closely related, yet computationally very distinct, identification scheme presented in Jeliazkov (2011), where the additive functions are identified through

$$\boldsymbol{y}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{M}_0\boldsymbol{Q}_{i2}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{M}_0\boldsymbol{Q}_{ir}\boldsymbol{g}_{ir} + \boldsymbol{\varepsilon}_i, \tag{11}$$

[3] It will be sufficient to apply this centering to $r-1$ of the unknown functions in each equation allowing the overall intercept to be absorbed in the remaining function.

where the $T \times T$ symmetric and idempotent mean-differencing matrix

$$\boldsymbol{M}_0 = \left(\boldsymbol{I}_T - \frac{\mathbf{1}_T \mathbf{1}_T'}{T} \right)$$

now pre-multiplies the incidence matrices $\{\boldsymbol{Q}_{ij}\}$ and centers the expanded vector of functional evaluations. The benefits from this identification method are discussed next.

3 Estimation

To motivate the general approach, this section begins by considering the important special case of a single equation univariate regression model. Given data $\{y_t, s_t\}_{t=1}^T$, the scalar responses y_t are assumed to depend on the (scalar) covariate s_t according to

$$y_t = g(s_t) + \varepsilon_t, \quad (t = 1, \ldots, T), \tag{12}$$

where $\varepsilon_t \sim N\left(0, \sigma^2\right)$, and $g(\cdot)$ is an unknown smooth function. The model in (12) can be written in stacked form as

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{g} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right), \tag{13}$$

where $\boldsymbol{Q}$ is the incidence matrix defined after equation (4). Given the Gaussian likelihood implied by (13), and assuming the Gaussian smoothness prior in (5) for either a first- or second-order process together with inverse Gamma priors $\tau^2 \sim IG(\nu_0/2, \delta_0/2)$ and $\sigma^2 \sim IG(s_0/2, d_0/2)$, yields full-conditional distributions which are conjugate, i.e. they are in the same family as the priors (see, e.g., Koop, 2003; Greenberg, 2008). Sequential sampling from those full-conditional distributions lays the foundations for the following algorithm.

Algorithm 1 ***Univariate Gaussian Nonparametric Model: MCMC Implementation***

1. *Sample* $[\boldsymbol{g}|\boldsymbol{y}, \tau^2, \sigma^2] \sim N(\hat{\boldsymbol{g}}, \boldsymbol{G})$*, where* $\boldsymbol{G}$ *and* $\hat{\boldsymbol{g}}$ *are the usual Bayes updates for linear regression, namely* $\boldsymbol{G} = \left(\boldsymbol{K}/\tau^2 + \boldsymbol{Q}'\boldsymbol{Q}/\sigma^2\right)^{-1}$ *and* $\hat{\boldsymbol{g}} = \boldsymbol{G}\left(\boldsymbol{K}\boldsymbol{g}_0/\tau^2 + \boldsymbol{Q}'\boldsymbol{y}/\sigma^2\right)$*. Remark 1 presents important notes on the sampling in this step.*
2. *Sample* $[\tau^2|\boldsymbol{g}] \sim IG\left(\frac{\nu_0+m}{2}, \frac{\delta_0+(\boldsymbol{g}-\boldsymbol{g}_0)'\boldsymbol{K}(\boldsymbol{g}-\boldsymbol{g}_0)}{2}\right)$*, where conditionally on* $\boldsymbol{g}$*,* τ^2 *is independent of the remaining parameters and the data.*

3. *Sample* $[\sigma^2|\boldsymbol{y},\boldsymbol{g}] \sim IG\left(\frac{s_0+n}{2}, \frac{d_0+(\boldsymbol{y}-\boldsymbol{Q}\boldsymbol{g})'(\boldsymbol{y}-\boldsymbol{Q}\boldsymbol{g})}{2}\right)$.

While steps 2 and 3 of Algorithm 1 are fairly straightforward, step 1 requires careful consideration because the quantities involved there can be of dimension as high as the sample size n. For this reason, estimation is performed as follows (see Fahrmeir and Lang, 2001).

Remark 1 ***Sampling of g.*** *To sample $\boldsymbol{g}$, note that $\boldsymbol{Q}'\boldsymbol{Q}$ is a diagonal matrix whose t-th diagonal entry equals the number of values in $\boldsymbol{s}$ corresponding to the design point $\boldsymbol{v}$. Since $\boldsymbol{K}$ and $\boldsymbol{Q}'\boldsymbol{Q}$ are banded, $\boldsymbol{G}^{-1}$ is banded as well. Thus sampling of $\boldsymbol{g}$ need not include an inversion to obtain $\boldsymbol{G}$ and $\hat{\boldsymbol{g}}$. The mean $\hat{\boldsymbol{g}}$ is found instead by solving $\boldsymbol{G}^{-1}\hat{\boldsymbol{g}} = \left(\boldsymbol{K}\boldsymbol{g}_0/\tau^2 + \boldsymbol{Q}'\boldsymbol{y}/\sigma^2\right)$, which is done in $O(T)$ operations by back substitution. Also, let $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{G}^{-1}$, where $\boldsymbol{P}$ is the Cholesky decomposition of $\boldsymbol{G}^{-1}$ and is also banded. To obtain a random draw from $N(\hat{\boldsymbol{g}},\boldsymbol{G})$ efficiently, sample $\boldsymbol{u} \sim N(\boldsymbol{0},\boldsymbol{I})$, and solve $\boldsymbol{P}\boldsymbol{w} = \boldsymbol{u}$ for $\boldsymbol{w}$ by back substitution. It follows that $\boldsymbol{w} \sim N(\boldsymbol{0},\boldsymbol{G})$. Adding the mean $\hat{\boldsymbol{g}}$ to $\boldsymbol{w}$, one obtains a draw $\boldsymbol{g} \sim N(\hat{\boldsymbol{g}},\boldsymbol{G})$.*

Turning attention to the additive case, let $\boldsymbol{\theta}$ denote the vector of all model parameters, i.e. the elements of $\{\boldsymbol{g}_{ij}\}$, $\{\tau_{ij}^2\}$, and the unique entries of $\boldsymbol{\Omega}$. Then, based on the identifying restrictions in (11) and the priors discussed in Section 2, MCMC estimation can proceed through iterative sampling of the following steps.

Algorithm 2 ***NPVAR Model: MCMC Implementation***

1. *Sample* $[\boldsymbol{g}_{i1}|\boldsymbol{y},\boldsymbol{\theta}\backslash\boldsymbol{g}_{i1}] \sim N\left(\hat{\boldsymbol{g}}_{i1},\hat{\boldsymbol{G}}_{i1}\right)$, *where,*

$$\hat{\boldsymbol{G}}_{i1} = \left(\frac{1}{\tau_{i1}^2}\boldsymbol{K}_{i1} + \frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{i1}'\boldsymbol{Q}_{i1}\right)^{-1},$$

$$\hat{\boldsymbol{g}}_{i1} = \hat{\boldsymbol{G}}_{i1}\left(\frac{1}{\tau_{i1}^2}\boldsymbol{K}_{i1}\boldsymbol{g}_{i10} + \frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{i1}'\left(\boldsymbol{y}_i - \mu_{i|\backslash i} - \sum_{j=2}^{r}\boldsymbol{M}_0\boldsymbol{Q}_{ij}\boldsymbol{g}_{ij}\right)\right),$$

with $\mu_{i|\backslash i} = E(\varepsilon_i|\boldsymbol{\varepsilon}_{\backslash i})$ and $\sigma_{i|\backslash i}^2 = Var(\varepsilon_i|\boldsymbol{\varepsilon}_{\backslash i})$. The sampling in this step is carried out efficiently in $O(T)$ operations as discussed in Remark 1.

2. *Sample* $[\boldsymbol{g}_{ij}|\boldsymbol{y},\boldsymbol{\theta}\backslash\boldsymbol{g}_{ij}] \sim N\left(\hat{\boldsymbol{g}}_j,\hat{\boldsymbol{G}}_j\right)$ *for $j = 2,\ldots,r$ and $i = 1,\ldots,q$, where*

$$\hat{\boldsymbol{G}}_{ij} = \left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij} + \frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{ij}'\boldsymbol{M}_0\boldsymbol{Q}_{ij}\right)^{-1}, \quad \textit{and}$$

$$\hat{\boldsymbol{g}}_{ij} = \hat{\boldsymbol{G}}_{ij}\left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij}\boldsymbol{g}_{ij0} + \frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{ij}'\boldsymbol{M}_0\left(\boldsymbol{y} - \mu_{i|\backslash i} - \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} - \sum_{k\geq 2, k\neq j}\boldsymbol{M}_0\boldsymbol{Q}_{ik}\boldsymbol{g}_{ik}\right)\right).$$

Remark 2 below shows how the sampling in this step can be carried out efficiently in $O(T)$ operations, even though $\hat{\boldsymbol{G}}_j$ is not banded.

3. *Sample* $[\tau_{ij}^2|\boldsymbol{g}_{ij}] \sim IG\left([\nu_{ij0}+m_j]/2, [\delta_{ij0}+\left(\boldsymbol{g}_{ij}-\boldsymbol{g}_{ij0}\right)'\boldsymbol{K}_{ij}\left(\boldsymbol{g}_{ij}-\boldsymbol{g}_{ij0}\right)]/2\right)$ *for* $i=1,\ldots,q$, *and* $j=1,\ldots,r$, *where, given* $\boldsymbol{g}_{ij}$, τ_{ij}^2 *is independent of the other elements in* $\boldsymbol{\theta}$ *and the data* $\boldsymbol{y}$.

4. *Sample* $[\boldsymbol{\Omega}^{-1}|\boldsymbol{y},\boldsymbol{\theta}\backslash\boldsymbol{\Omega}] \sim W\left(r_0+T, [R_0^{-1}+\sum_{t=1}^{T}\boldsymbol{e}_t\boldsymbol{e}_t']^{-1}\right)$, *where* $\boldsymbol{e}_t$ *denotes the* $q\times 1$ *vector of residuals in time period* t.

Algorithm 2 generalizes Algorithm 1 in a straightforward fashion by sampling each unknown function conditionally on the remaining ones, making simulation manageable. Importantly, however, Step 2 of Algorithm 2 involves $r-1$ non-banded matrices in each equation, and at first glance it would appear that simulation will be very demanding. Fortunately, however, as shown in Jeliazkov (2011), an application of the Sherman-Morrison formula makes it possible to sample these functions efficiently. The approach is presented in greater detail in Remark 2 further below. The modularity and computational advantages of this estimation strategy can provide important benefits in a variety of settings because simulating the functions by brute force methods is not always practical owing to the algorithmic complexity of working with high-dimensional matrices. Moreover, because the frequentist backfitting approach to estimating the unknown functions can be viewed as a (non-stochastic) simplification of Gibbs sampling (see Hastie and Tibshirani, 2000), Algorithm 2 can also be useful in frequentist estimation.

Remark 2 ***Sampling of Centered Functions.*** *To draw* $\boldsymbol{g}_{ij} \sim N\left(\hat{\boldsymbol{g}}_{ij}, \hat{\boldsymbol{G}}_{ij}\right)$ *in Step 2 of Algorithm 2, use the definition of* $\boldsymbol{M}_0$ *to write*

$$\begin{aligned}\hat{\boldsymbol{G}}_{ij} &= \left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij}+\frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{ij}'\boldsymbol{M}_0\boldsymbol{Q}_{ij}\right)^{-1}\\ &= \left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij}+\frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{ij}'\boldsymbol{Q}_{ij}-\frac{\boldsymbol{c}_{ij}\boldsymbol{c}_{ij}'}{\sigma_{i|\backslash i}^2 T}\right)^{-1},\end{aligned}$$

where $\boldsymbol{c}_{ij}=\boldsymbol{Q}_{ij}'\boldsymbol{1}$. *Letting* $\boldsymbol{A}_{ij}=\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij}+\frac{1}{\sigma_{i|\backslash i}^2}\boldsymbol{Q}_{ij}'\boldsymbol{Q}_{ij}$, $\boldsymbol{u}_{ij}=\frac{1}{\sqrt{\sigma_{i|\backslash i}^2 T}}\boldsymbol{c}_{ij}$, *and* $\lambda_{ij}=\boldsymbol{u}_{ij}'\boldsymbol{A}_{ij}^{-1}\boldsymbol{u}_{ij}$, *one can write, by the Sherman-Morrison formula,*

$$\begin{aligned}\hat{\boldsymbol{G}}_{ij} &= \left(\boldsymbol{A}_{ij}-\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'\right)^{-1}\\ &= \boldsymbol{A}_{ij}^{-1}+\frac{\boldsymbol{A}_{ij}^{-1}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'\boldsymbol{A}_{ij}^{-1}}{1-\lambda_{ij}}.\end{aligned} \tag{14}$$

Significant efficiency benefits can be derived from (14) *because* $\hat{\boldsymbol{g}}_j$ *in Step 2 of Algorithm 2 can be obtained by working with* $\boldsymbol{A}_{ij}$ *without inverting to* $\boldsymbol{A}_{ij}^{-1}$ *as outlined in Remark 1. Furthermore, let*

$$\boldsymbol{B}_{ij}=\left(\boldsymbol{A}_{ij}+\frac{\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}'}{1-\lambda_{ij}}\right),$$

which implies that $\hat{\boldsymbol{G}}_{ij} = \boldsymbol{A}_{ij}^{-1}\boldsymbol{B}_{ij}\boldsymbol{A}_{ij}^{-1}$. *Thus, if* $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{B}_{ij})$, *then* $\boldsymbol{z} = \boldsymbol{A}_{ij}^{-1}\boldsymbol{x}$ *is distributed* $\boldsymbol{z} \sim N\left(\boldsymbol{0}, \hat{\boldsymbol{G}}_{ij}\right)$, *and a draw* $\boldsymbol{g}_{ij} \sim N\left(\hat{\boldsymbol{g}}_{ij}, \hat{\boldsymbol{G}}_{ij}\right)$ *is obtained as* $\boldsymbol{g}_{ij} = \hat{\boldsymbol{g}}_{ij} + \boldsymbol{z}$. *To generate* $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{B}_{ij})$, *draw* $\boldsymbol{w}_1 \sim N(\boldsymbol{0}, \boldsymbol{A}_{ij})$ *and* $w_2 \sim N(0, 1)$ *and let* $\boldsymbol{x} = \boldsymbol{w}_1 + w_2\boldsymbol{u}_{ij}/\sqrt{1 - \lambda_{ij}}$,

As a consequence of the shortcuts afforded by Remark 2, all operations are $O(T)$ rather than $O\left(T^3\right)$.

4 Model Comparison

Empirical studies must inevitably address uncertainty not only about the parameters of a given model, but also about the model specification itself. This makes model comparison a central issue in statistical analysis. Given a collection of models $\{\mathcal{M}_1, \ldots, \mathcal{M}_L\}$, the formal Bayesian approach to model comparison (or testing the validity of the alternative hypotheses captured by each model) is based on the posterior model probabilities and their ratios, the posterior odds. Specifically, for any two models $\mathcal{M}_i$ and $\mathcal{M}_j$, a simple application of Bayes' theorem suggests that the posterior odds can be represented as the product of the prior odds and the ratio of the marginal likelihoods (the Bayes factor) as follows

$$\frac{\Pr(\mathcal{M}_i|\boldsymbol{y})}{\Pr(\mathcal{M}_j|\boldsymbol{y})} = \frac{\Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j)} \times \frac{m(\boldsymbol{y}|\mathcal{M}_i)}{m(\boldsymbol{y}|\mathcal{M}_j)}.$$

In turn, for any model $\mathcal{M}_l$, $l = 1, \ldots, L$, the marginal likelihood is given by

$$m(\boldsymbol{y}|\mathcal{M}_l) = \int f(\boldsymbol{y}|\boldsymbol{\theta}_l, \mathcal{M}_l)\pi_l(\boldsymbol{\theta}_l|\mathcal{M}_l)d\boldsymbol{\theta}_l, \tag{15}$$

which is the integral of the likelihood function $f(\boldsymbol{y}|\boldsymbol{\theta}_l, \mathcal{M}_l)$ with respect to the prior distribution on the model parameters $\pi(\boldsymbol{\theta}_l|\mathcal{M}_l)$. Because in the case of nonparametric additive models the dimension of $\boldsymbol{\theta}$ can be very large, it should be clear that direct analytical integration will generally be infeasible. However, this difficulty can be addressed by using the approach of Chib (1995), where after rearranging Bayes' theorem $m(\boldsymbol{y}|\mathcal{M}_l)$ can alternatively be expressed as

$$m(\boldsymbol{y}|\mathcal{M}_l) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}_l^*, \mathcal{M}_l)\pi(\boldsymbol{\theta}_l^*|\mathcal{M}_l)}{\pi(\boldsymbol{\theta}_l^*|\boldsymbol{y}, \mathcal{M}_l)}, \tag{16}$$

so that the integral in (15) is reduced to the more tractable problem of evaluating the likelihood, prior, and posterior ordinates at a single point $\boldsymbol{\theta}_l^*$ (e.g., the posterior mean). Because the numerator terms in (16)

are available by direct calculation, the marginal likelihood can be computed by finding an estimate of the posterior ordinate $\pi(\boldsymbol{\theta}^*|\boldsymbol{y})$.

In the current context, the hierarchical structure of NPVAR models allows application of (16) in two different ways. One approach relies on

$$m(\boldsymbol{y}) = \frac{f\left(\boldsymbol{y}|\left\{\tau_{ij}^{2*}\right\}, \boldsymbol{\Omega}^*, \left\{\boldsymbol{g}_{ij}^*\right\}\right) \pi\left(\left\{\tau_{ij}^{2*}\right\}, \boldsymbol{\Omega}^*, \left\{\boldsymbol{g}_{ij}^*\right\}\right)}{\pi\left(\left\{\tau_{ij}^{2*}\right\}, \boldsymbol{\Omega}^*, \left\{\boldsymbol{g}_{ij}^*\right\}|\boldsymbol{y}\right)},$$

where the nonparametric functions are explicitly included in the identity (see Chib and Jeliazkov, 2006; Chib et al., 2009). However, owing to the Gaussian structure of the model, the marginal likelihood can also be computed using

$$m(\boldsymbol{y}) = \frac{f\left(\boldsymbol{y}|\{\tau_{ij}^{2*}\}, \boldsymbol{\Omega}^*\right) \pi\left(\{\tau_{ij}^{2*}\}, \boldsymbol{\Omega}^*\right)}{\pi\left(\{\tau_j^{2*}\}, \boldsymbol{\Omega}^*|\boldsymbol{y}\right)},$$

where all quantities are marginalized over the high-dimensional blocks $\{\boldsymbol{g}_{ij}\}$. This marginalization is possible because conditionally on $\left(\left\{\tau_{ij}^{2*}\right\}, \boldsymbol{\Omega}^*\right)$, the density $f\left(\boldsymbol{y}|\{\tau_{ij}^{2*}\}, \boldsymbol{\Omega}^*\right)$, marginalized over $\{\boldsymbol{g}_{ij}\}$ with respect to the prior distributions in (5), is also normal (Koop and Poirier, 2004) and can be evaluated directly for the typical sample sizes T encountered in macroeconomic applications. Because of this analytical tractability, $m\left(\boldsymbol{y}\right)$ can then be found after the main run where, using the conditional independence of the densities in Steps 3 and 4 of Algorithm 2, one computes

$$\pi\left(\left\{\tau_{ij}^{2*}\right\}, \boldsymbol{\Omega}^*|y\right) \approx T^{-1} \sum_{t=1}^{T} \left\{ f_{IW}\left(\boldsymbol{\Omega}^*|\boldsymbol{y}, \left\{\boldsymbol{g}_j^{(t)}\right\}\right) \prod_{i=1}^{p} f_{IG}\left(\tau_j^{2*}|\boldsymbol{g}_j^{(t)}\right) \right\}$$

using draws $\{\boldsymbol{g}_j^{(t)}\}$ from the main MCMC run. In instances where T is large or there are other complications (e.g., discrete outcomes), the decomposition involving $\{\boldsymbol{g}_{ij}\}$ is more appropriate and readers are referred to Chib and Jeliazkov (2006) and Chib et al. (2009) for methods that use reduced runs or to Jeliazkov and Lee (2010) for a method that employs the Gibbs kernel and invariance of the Markov chain to estimate the posterior ordinate $\pi(\{\tau_{ij}^{2*}\}, \boldsymbol{\Omega}^*, \{\boldsymbol{g}_{ij}^*\}|\boldsymbol{y})$.

An important point mentioned earlier is that the marginal likelihood for the additive NPVAR model does not depend on the (likelihood unidentified) levels of the functions that are centered for identification. This

can be seen by recognizing that if $f(\boldsymbol{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = f(\boldsymbol{y}|\boldsymbol{\theta}_1)$, i.e. the likelihood depends only on $\boldsymbol{\theta}_1$ whereas $\boldsymbol{\theta}_2$ is unidentified, but we have a proper prior $\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, then the marginal likelihood

$$\begin{aligned} m(\boldsymbol{y}) &= \int f(\boldsymbol{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\, \pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\, d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \\ &= \int f(\boldsymbol{y}|\boldsymbol{\theta}_1)\, \pi(\boldsymbol{\theta}_1) \int \pi(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)\, d\boldsymbol{\theta}_2 d\boldsymbol{\theta}_1 \\ &= \int f(\boldsymbol{y}|\boldsymbol{\theta}_1)\, \pi(\boldsymbol{\theta}_1)\, d\boldsymbol{\theta}_1, \end{aligned}$$

is not influenced by the prior on $\boldsymbol{\theta}_2$. In practice this is important for modeling, because it implies that researchers with different beliefs about unidentified parameters will nevertheless reach identical conclusions about the relative ranking of alternative models.

Finally, note that because estimation of the marginal likelihood does not require maximization, it is less computationally intensive in nonparametric additive models than evaluation of information criteria such as AIC and BIC. This point has been overlooked and not fully appreciated in the literature despite its importance for model selection and model averaging on the basis of $\{\Pr(\mathcal{M}_l|\boldsymbol{y})\}$.

5 Model Extensions

The estimation techniques presented in this paper are fully modular and readily applicable in various other settings since estimation of the unknown functions $\{\boldsymbol{g}_{ij}\}$ can be done conditionally on modifications in other parts of the model. The goal of this discussion is to briefly review the relevant literature and provide references that could guide researchers interested in pursuing such extensions.

One should note that the methods in Section 2 and Section 3 trivially generalize to cases where one or more exogenous covariates enter the regression as in seemingly unrelated regression models (e.g., Smith and Kohn, 2000; Holmes et al., 2002; Koop et al., 2005). Further extensions of the framework to Bayesian models with nonparametric endogeneity or sample selection can be pursued following Chib and Greenberg (2007) or Chib et al. (2009), respectively; the modeling would also be useful in guiding future research on structural NPVAR models and impulse response analysis. Many of the aforementioned papers also trivially subsume semiparametric and partially linear cases where some of the covariates enter the model linearly.

Estimation of such models is a straightforward extension of Algorithms 1 and 2, and proceeds by using the partial residuals $\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$ when simulating $\{\boldsymbol{g}_{ij}\}$, followed by simulating $\boldsymbol{\beta}$ conditionally upon the functions $\{\boldsymbol{g}_{ij}\}$.

The methods in this paper, including the above extensions, can also be applied, using data augmentation techniques (Tanner and Wong, 1987; Albert and Chib, 1993), to the analysis of dynamic systems involving binary, polychotomous, censored, and other discrete outcomes such as the qualitative VAR model of Dueker (2005). The main advantage of this approach is that conditionally on the latent data, estimation of the parameters and the unknown functions closely mirrors the methods for continuous data. The methodology presented here is also applicable to the class of additive mixed models for continuous and discrete data (e.g., Lin and Zhang, 1999). For example, Chib and Jeliazkov (2006) discuss the specification and estimation of a semiparametric partially linear model for dynamic binary panel data with multivariate heterogeneity. The estimation algorithm in that paper can be easily modified to include an additive structure whose estimation can be carried out by the methods presented in Section 3.

While the specification and estimation of NPVAR models was discussed in detail for homoskedastic Gaussian models, extensions to other distributions (e.g., Student's t, mixtures of normals, or Dirichlet process priors for nonparametric distributional modeling) and heteroskedasticity (e.g., regime switching models with different variance regimes, or models with stochastic volatility) may be very desirable in certain applications. One such example, relating to different variance regimes, will be studied in Section 6. Fortunately, such extensions can be estimated using data augmentation techniques that could build upon the homoskedastic Gaussian specification discussed earlier. In particular, consider a heteroskedastic model in which the ith equation can be written as

$$\boldsymbol{y}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{M}_0\boldsymbol{Q}_{i2}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{M}_0\boldsymbol{Q}_{ir}\boldsymbol{g}_{ir} + \boldsymbol{\varepsilon}_i$$

with $\boldsymbol{\Sigma}_i \equiv \mathrm{Var}(\boldsymbol{\varepsilon}_\mathrm{i}) = \mathrm{diag}(\sigma^2_{\mathrm{i1}}, \ldots, \sigma^2_{\mathrm{iT}})$. Due to the heteroskedasticity, the covariance matrix of $[\boldsymbol{g}_{ij}|\boldsymbol{y}, \boldsymbol{\theta}]$ is not of the form presented in Remark 2, and estimation can not be performed efficiently by relying on the Sherman-Morrison formula. This poses an important computational difficulty because estimation in large

dimensional models would be very difficult and potentially infeasible. In this paper, I propose a solution to this problem that employs data augmentation to reduce the heteroskedastic model to a homoskedastic one enabling application of the methods discussed in Algorithm 2 and Remark 2. In particular, following Chib and Jeliazkov (2006), we can write

$$\boldsymbol{y}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{M}_0\boldsymbol{Q}_{i2}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{M}_0\boldsymbol{Q}_{ir}\boldsymbol{g}_{ir} + \boldsymbol{\eta}_i + \boldsymbol{\nu}_i,$$

where $\boldsymbol{\eta}_i \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_i - \kappa_i\boldsymbol{I})$ and $\boldsymbol{\nu}_i \overset{iid}{\sim} N(\mathbf{0}, \kappa_i\boldsymbol{I})$ with $0 < \kappa_i \leq \min\{\sigma_{it}^2\}$. Consequently, given a draw of $\boldsymbol{\eta}_i$, which is simple and inexpensive to obtain, the model

$$\boldsymbol{y}_i - \boldsymbol{\eta}_i = \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \boldsymbol{M}_0\boldsymbol{Q}_{i2}\boldsymbol{g}_{i2} + \ldots + \boldsymbol{M}_0\boldsymbol{Q}_{ir}\boldsymbol{g}_{ir} + \boldsymbol{\nu}_i$$

is homoskedastic because $\mathrm{Var}(\boldsymbol{\nu}_\mathrm{i}) = \kappa_\mathrm{i}\boldsymbol{I}$. In our context, it would actually be optimal to set $\kappa_i = \min\{\sigma_{it}^2\}$ because this would imply that the corresponding elements of $\boldsymbol{\eta}_i$ would be identically 0 and will not need to be sampled. This leads to the following extension of Algorithm 2:

Algorithm 3 ***NPVAR Model: MCMC Estimation of Heteroskedastic Model***

1. *For $i = 1, \ldots, q$:*

   (a) *Sample $[\boldsymbol{\eta_i}|\boldsymbol{y}, \boldsymbol{\theta}]$ by drawing, for $t = 1, \ldots, T$, $\eta_{it} \sim N(\hat{\eta}_{it}, \hat{H}_{it})$, where $\hat{H}_{it} = \kappa_i(\sigma_{it}^2 - \kappa_i)/\sigma_{it}^2$ and $\hat{\eta}_{it} = (\sigma_{it}^2 - \kappa_i)(y_{it} - \mu_{it|\backslash i,t} - m_{it})/\sigma_{it}^2$, where m_{it} is the t-th row of $\boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} + \sum_{k\geq 2}\boldsymbol{M}_0\boldsymbol{Q}_{ik}\boldsymbol{g}_{ik}$, $\mu_{it|\backslash i,t}$ is the t-th row of $\boldsymbol{\mu}_{i|\backslash i} = E(\boldsymbol{\varepsilon}_i|\boldsymbol{\varepsilon}_{\backslash i})$, and $\kappa_i = \min\{\sigma_{it}^2\}$, where $\sigma_{it}^2 = Var(\boldsymbol{\varepsilon}_{it}|\boldsymbol{\varepsilon}_{\backslash i,t})$. Note that for cases where $\kappa_i = \sigma_{it}^2$, the corresponding entry in $\boldsymbol{\eta}_i$ is identically zero and need not be sampled.*

   (b) *Sample $[\boldsymbol{g}_{i1}|\boldsymbol{y}, \boldsymbol{\eta}_i, \boldsymbol{\theta}\backslash\boldsymbol{g}_{i1}] \sim N\left(\hat{\boldsymbol{g}}_{i1}, \hat{\boldsymbol{G}}_{i1}\right)$, where,*

$$\hat{\boldsymbol{G}}_{i1} = \left(\frac{1}{\tau_{i1}^2}\boldsymbol{K}_{i1} + \frac{1}{\kappa_i}\boldsymbol{Q}_{i1}'\boldsymbol{Q}_{i1}\right)^{-1},$$

$$\hat{\boldsymbol{g}}_{i1} = \hat{\boldsymbol{G}}_{i1}\left(\frac{1}{\tau_{i1}^2}\boldsymbol{K}_{i1}\boldsymbol{g}_{i10} + \frac{1}{\kappa_i}\boldsymbol{Q}_{i1}'\left(\boldsymbol{y}_i - \boldsymbol{\eta}_i - \boldsymbol{\mu}_{i|\backslash i} - \sum_{j=2}^{r}\boldsymbol{M}_0\boldsymbol{Q}_{ij}\boldsymbol{g}_{ij}\right)\right),$$

   with $\boldsymbol{\mu}_{i|\backslash i} = E(\boldsymbol{\varepsilon}_i|\boldsymbol{\varepsilon}_{\backslash i})$ and $\kappa_i = \min\{\sigma_{it}^2\}$, where $\sigma_{it}^2 = Var(\boldsymbol{\varepsilon}_{it}|\boldsymbol{\varepsilon}_{\backslash i,t})$. The sampling in this step is carried out efficiently as in Remark 1.

   (c) *Sample $[\boldsymbol{g}_{ij}|\boldsymbol{y}, \boldsymbol{\eta}_i, \boldsymbol{\theta}\backslash\boldsymbol{g}_{ij}] \sim N\left(\hat{\boldsymbol{g}}_j, \hat{\boldsymbol{G}}_j\right)$ for $j = 2, \ldots, r$ and $i = 1, \ldots, q$, where*

$$\hat{\boldsymbol{G}}_{ij} = \left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij} + \frac{1}{\kappa_i}\boldsymbol{Q}_{ij}'\boldsymbol{M}_0\boldsymbol{Q}_{ij}\right)^{-1}, \quad \textit{and}$$

$$\hat{\boldsymbol{g}}_{ij} = \hat{\boldsymbol{G}}_{ij}\left(\frac{1}{\tau_{ij}^2}\boldsymbol{K}_{ij}\boldsymbol{g}_{ij0} + \frac{1}{\kappa_i}\boldsymbol{Q}_{ij}'\boldsymbol{M}_0\left(\boldsymbol{y} - \boldsymbol{\eta}_i - \boldsymbol{\mu}_{i|\backslash i} - \boldsymbol{Q}_{i1}\boldsymbol{g}_{i1} - \sum_{k\geq 2, k\neq j}\boldsymbol{M}_0\boldsymbol{Q}_{ik}\boldsymbol{g}_{ik}\right)\right).$$

This is done as in Remark 2.

2. *Sample* $[\tau_{ij}^2|\boldsymbol{g}_{ij}]$ *for* $i = 1, \ldots, q$, *and* $j = 1, \ldots, r$, *as in Algorithm 2.*

3. *Sample* $[\boldsymbol{\Omega_{it}}^{-1}|\boldsymbol{y}, \boldsymbol{\theta}\backslash\boldsymbol{\Omega}]$ *according to the volatility process under consideration.*

The above machinery also applies to mixture-of-normals and scale mixture-of-normals models. In particular, a model with t errors with ν degrees of freedom can be represented as a conditionally Gaussian model, whose variance, given a set of *a priori* gamma latent variables $\lambda_t \sim G(\nu/2, \nu/2)$, $t = 1, \ldots, T$, is given by $\mathrm{Var}(\varepsilon_{\mathrm{t}}|\lambda_{\mathrm{t}}) = \sigma^2/\lambda_{\mathrm{t}}$ (Andrews and Mallows, 1974; Albert and Chib, 1993). Estimation of these models is straightforward because given $\{\lambda_i\}$, one can decompose $\boldsymbol{\varepsilon}_i$ into $\boldsymbol{\eta}_i$ and $\boldsymbol{\nu}_i$ and proceed as above. In this way, NPVAR models can be adapted to a variety of specifications for the error variance, including changepoint and regime switching models (Chib, 1996, 1998; Sims and Zha, 2006), time-varying parameter models (Primiceri, 2005; Chan and Jeliazkov, 2009), factor models (Kose et al., 2003; Belviso and Milani, 2006; Kose et al., 2008; Chan and Jeliazkov, 2009), and others.

6 Application to U.S. Macroeconomic Data

The data sample for this application contains post-war quarterly macroeconomic data for the U.S. from 1948:Q1 to 2005:Q1. The set of variables includes output growth g_t measured by log differences of real GDP between two consecutive quarters, average quarterly unemployment rate u_t, inflation 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 first three of these variables are seasonally adjusted. These variables, summarized in Table 1, reflect the general state of the economy, and have been widely used in empirical macroeconomics.[4] From the Table, we see that

[4]The sample period excludes the past recession for a number of reasons. Over the last few years, interest rates have approached and stayed very close to their lower bound of zero. This could lead to findings of nonlinearity due to the effects of the lower bound, thereby favoring the methods of the paper over a linear model. Moreover, traditional modeling may be inadequate near the bound, where the distribution of the interest rate process is truncated and exhibits point mass. Appropriate modeling in this case is still an open research problem. Finally, if the "Great Recession" marked a possible structural break, at present there would be insufficient observations estimate the model after the break.

the average quarterly GDP growth over the sample period is 0.85 percent, which amounts to annual GDP growth of 3.4 percent. A similar computation shows an average annual inflation rate of approximately 3.7 percent. Unemployment and interest rates average at 5.63 and 4.81 percent, respectively.

Table 1: Descriptive statistics for the data sample (in percentage points).

Variable	Mean	SD	Min	Max
Quarterly growth in real GDP	0.85	1.00	-2.76	4.02
Unemployment rate	5.63	1.52	2.60	10.70
Nominal interest rate	4.81	2.92	0.79	15.05
Quarterly Inflation	0.92	0.85	-1.24	4.08

These data are analyzed using the econometric techniques discussed earlier. The empirical strategy for studying the behavior of the dynamic system in (2) is to address both model and functional form uncertainty. The first area of model uncertainty in the macroeconomic system has to do with determination its dynamics – i.e., the number of lags needed in equation (2). This one-lag model was compared to several more richly parameterized models in order to gauge whether restricting attention to an NPVAR(1) specification is a sensible empirical strategy. The baseline NPVAR(1) model, which contains a single lag of $\boldsymbol{y}_t$ with 16 unknown functions, was compared with an NPVAR(2) model (with 32 such functions). The baseline model overwhelmingly outperformed the longer lag specification – its log-marginal likelihood exceeded that of the larger model by over 40, implying a Bayes factor of over e^{40} in favor of the NPVAR(1) specification.[5]

Guided by earlier research findings suggesting the possibility of a structural break (at least in error volatility as in Stock and Watson (2003) and Sims and Zha (2006)), I also used split-sample estimation to capture the possibility of structural breaks in the series. Specifically, an NPVAR(1) model was fit on data in the pre-Volcker era (prior to 1979:Q2), and a separate model was fit on the data thereafter (following 1979:Q3). The marginal likelihood for the pre-Volcker model was -539.5, and that for the second part of the data sample was -438.7. Compared to the log-marginal likelihood of -908.8 for the baseline NPVAR(1) model on the entire data sample, the split sample measure of fit, as captured by the marginal likelihood, was

[5] A specification including a fourth (year-ago) lag was also considered but it also did not perform competitively with the NPVAR(1) specification.

far worse (the sum of the log-marginal likelihoods for the two subsamples is -978.2, which is far below marginal likelihood for the overall NPVAR(1) model of -908.8). These results are interesting because (i) they suggest that the simpler and parsimonious NPVAR(1) fit on the entire specification appears preferable to the (twice as big) split sample model and (ii) they demonstrate the ability of the Bayesian model comparison framework to penalize overparameterized specifications.

Figure 2: Full sample estimates: the rows represent the functions in each equation, columns contain the functions of a given lagged variable across equations.

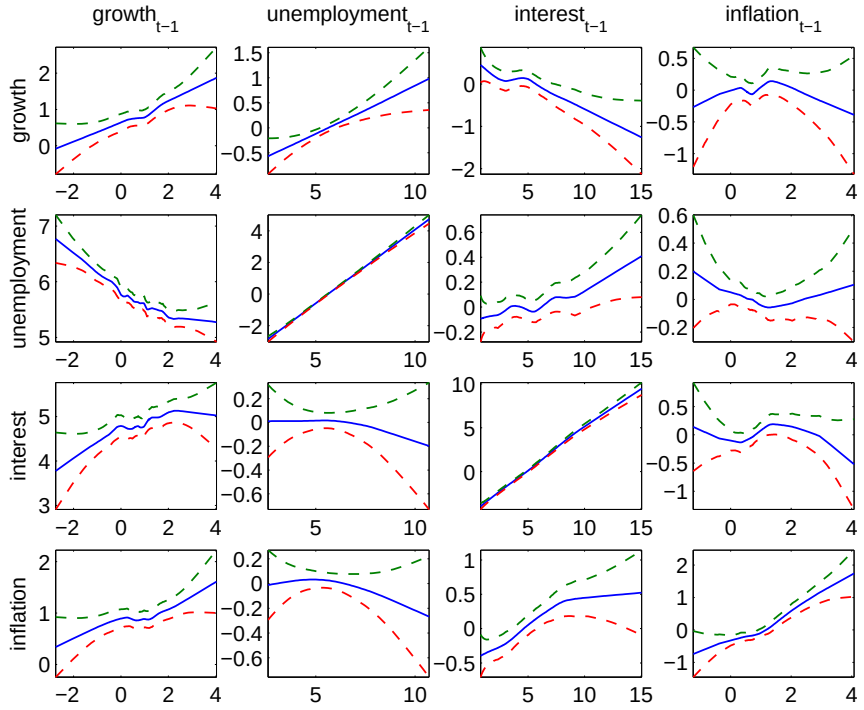


Figure 2 presents the estimated functions for the full-sample NPVAR(1) model. The figure shows that a linear model would be reasonable for many of the economic relationships – particularly in modeling the effects of lagged unemployment. To a lesser extent, the same is true in other instances (e.g. the dependence of interest on its past value), where the function estimates do not reveal drastic departures from linearity. On the other hand, however, in many equations, the effects of lagged financial variables (interest and inflation), as well as the effects of lagged growth, appear to be quite nonlinear. This finding concurs with earlier studies that have found nonlinearity in growth behavior (Dahl and Gonzalez-Rivera, 2003a,b) and financial

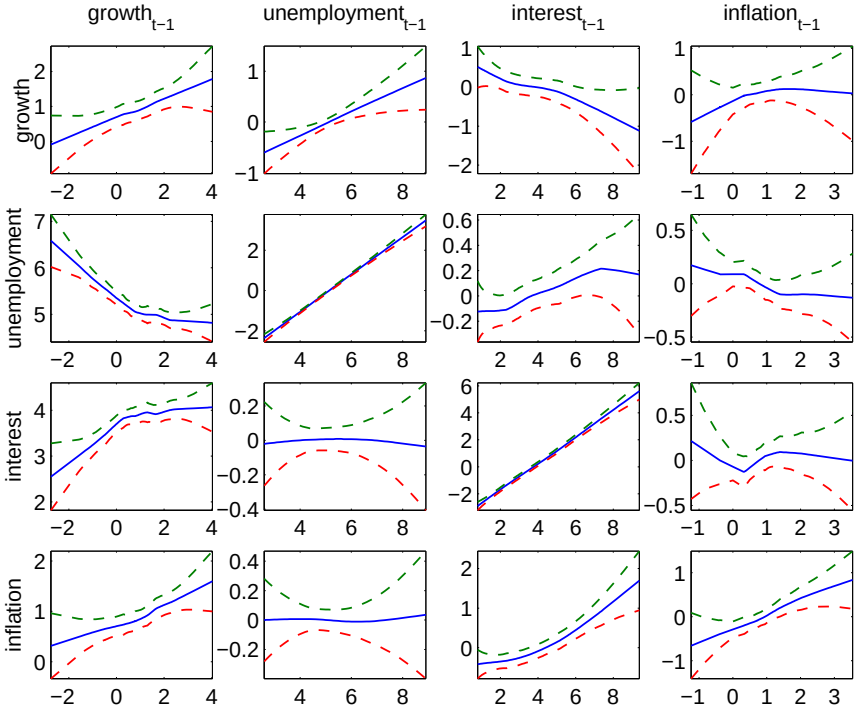
markets (Härdle and Tsybakov, 1997; Härdle et al., 1998). The Figure shows, for instance, that lagged inflation exhibits significant nonlinearities in every equation, whereas the function estimates for lagged interest exhibit nonlinearities in three of the four equations. For this reason, future analysis of financial variables might benefit considerably from employing nonparametric methods. Regarding the effects of lagged GDP growth, a review of the estimated functions reveals that although there is much nonlinearity, there are also large regions where the function estimates are approximately linear. This suggests that an interesting future research question would be to examine whether those types of nonlinearities can be adequately captured through threshold models.

For comparison purposes, Figures 3 and 4 show function estimates for the pre- and post-Volcker periods, respectively. It is interesting to note that, although the function estimates differ in some respects, most point to the same types of nonlinearities as the estimates from the overall sample.[6] This is quite instructive, as it provides evidence that the econometric relationships may be stable but nonlinear, and therefore omitted nonlinearity may be a significant driver in findings of structural instability (cf. Hamilton, 2001). Resolving this issue should be an important item on the research agenda of studies focusing on structural (in)stability.

The apparent stability of the nonparametric function estimates across subsamples naturally leads to another important research question that has attracted much attention recently. Specifically, it would be of interest to consider whether an NPVAR model would exhibit evidence of a structural change in variances, which has been widely documented in contexts utilizing linear models. Such findings (e.g., Stock and Watson (1996, 2003), Sims and Zha (2006)) have led to the conclusion that a reduction in error volatility has been a driving force in the "Great Moderation" of the 1980s and 1990s. To formally test the stability of the mean relationships while allowing for structural breaks in variances, I have estimated three additional NPVAR models. The first allows for a single structural break between 1979:Q2 and 1979:Q3 with the Volcker appointment. The second model employs a single structural break between 1982:Q4 and 1983:Q1 with the following the Fed's disinflation of the early 1980s. The third model allows for both of these break points. The models were estimated using Algorithm 3 of Section 5, and the marginal likelihoods

[6] Since the range and level of each function may differ across samples, readers are cautioned to compare those functions over the relevant ranges, keeping in mind that the level of the functions will shift to satisfy the identification constraints.

Figure 3: Pre-Volcker estimates: the rows represent the functions in each equation, columns contain the functions of a given lagged variable across equations.

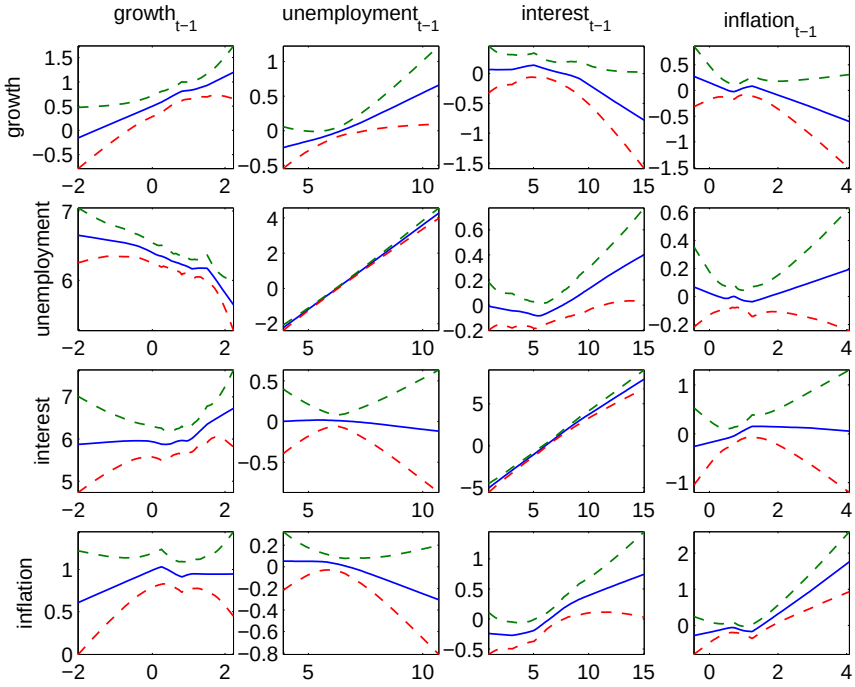


were estimated as discussed in Section 4. The marginal likelihood for the first model was estimated to be -891.4, whereas that of the second was estimated to be -880.1, showing that, conditionally on a single break, the data favor the 1982/83 breakpoint. However, a much more dramatic improvement is offered by the third model, the one which allows for both a 1979 and a 1982/83 breakpoints. The marginal likelihood for that model is -838.4, leading to the conclusion that these three periods in the U.S. sample are indeed dramatically different. This if further confirmed by examining the estimated covariance matrices for the three sub-periods:

$$\mathbf{\Omega}_{48:79} = \begin{pmatrix} 1.135 & -0.254 & 0.085 & -0.060 \\ -0.254 & 0.144 & -0.033 & 0.002 \\ 0.085 & -0.033 & 0.255 & 0.044 \\ -0.060 & 0.002 & 0.044 & 0.463 \end{pmatrix},$$

$$\mathbf{\Omega}_{79:82} = \begin{pmatrix} 1.327 & -0.285 & 1.095 & 0.636 \\ -0.285 & 0.367 & -0.575 & -0.261 \\ 1.095 & -0.575 & 4.348 & 1.445 \\ 0.636 & -0.261 & 1.445 & 1.048 \end{pmatrix},$$

Figure 4: Post-Volcker estimates: the rows represent the functions in each equation, columns contain the functions of a given lagged variable across equations.



and

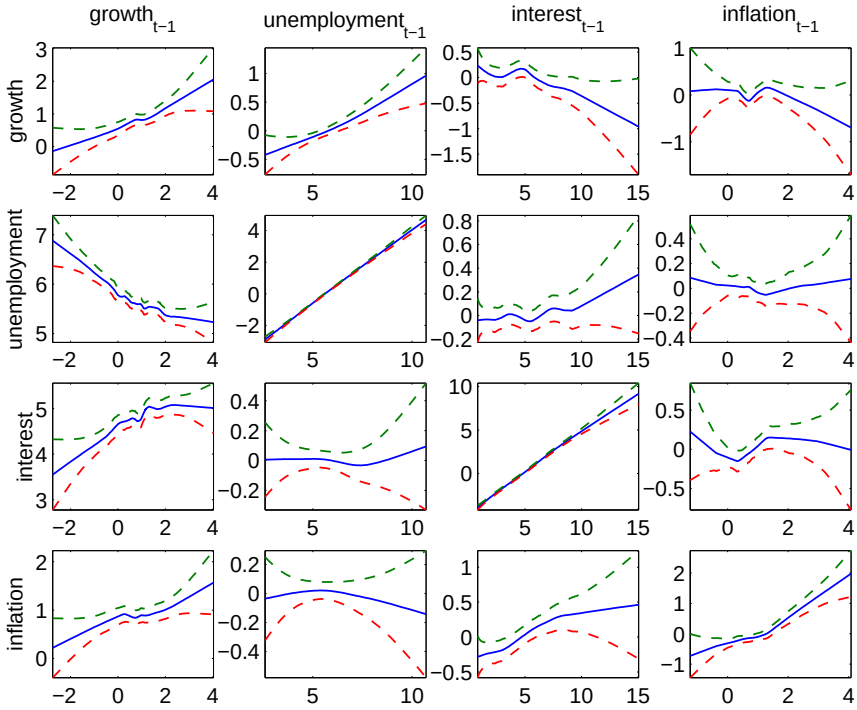
$$\mathbf{\Omega}_{83:05} = \begin{pmatrix} 0.234 & -0.032 & 0.051 & 0.008 \\ -0.032 & 0.066 & -0.031 & -0.010 \\ 0.051 & -0.031 & 0.202 & 0.044 \\ 0.008 & -0.010 & 0.044 & 0.216 \end{pmatrix}.$$

These covariance matrices clearly demonstrate the dramatic peak in the error variances of all variables except growth (i.e. unemployment, interest rates, and inflation) during the disinflation period and the subsequent "moderation" of all 4 variables in 1983. A notable feature is the large jump, and subsequent decrease, in the estimated error variance in the interest rate equation during the period 1979-1982, which can be accounted for by the Fed's change of policy instrument from the federal funds rate to reserve targeting, as well as the unprecedented increase in interest rates during the disinflation period.

Figure 5 presents the function estimates from the model with three variance regimes. The figure demonstrates that the same type of nonlinearities that were present in the homoskedastic models are still present here. Therefore, even though the heteroskedastic NPVAR model has confirmed earlier conclusions that

changes were large due to breaks in variances, it also shows that there is much nonlinearity that would remain unexplored by linear models and that future research should study such features of the economic relationships more closely.

Figure 5: Full sample estimates from model with 3 volatility regimes: the rows represent the functions in each equation, columns contain the functions of a given lagged variable across equations.



7 Concluding Remarks

This article has examined the specification, estimation, and comparison of nonparametric VAR models. Efficient MCMC sampling and model comparison techniques are discussed in the context of a new scheme for identifying the unknown covariate functions, and extensions to heteroskedastic and other settings have been examined. An application of the NPVAR model to U.S. post-war data on GDP growth, unemployment, interest rates, and inflation, has confirmed the presence of distinct volatility regimes in the post-war U.S. macroeconomic series, but has also revealed that important nonlinearities is certain economic relationships may remain undetected by standard regressions. Implementation of these techniques in related settings, such

as those considered in Section 5, is an interesting area for future research.

References

Ahamada, I. and Flachaire, E. (2010), *Non-Parametric Econometrics*, Oxford: Oxford University Press.

Albert, J. and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Andrews, D. F. and Mallows, C. L. (1974), "Scale Mixtures of Normal Distributions," *Journal of the Royal Statistical Society – Series B*, 36, 99–102.

Beaudry, P. and Koop, G. (1993), "Do Recessions Permanently Change Output?" *Journal of Monetary Economics*, 31, 149–163.

Belviso, F. and Milani, F. (2006), "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy," *Topics in Macroeconomics*, 6, Iss. 3, Article 2.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems," *Statistical Science*, 10, 3–66.

Canova, F. (1993), "Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model," *Journal of Economic Dynamics and Control*, 17, 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, 101–120.

Chauvet, M. (1998), "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching," *International Economic Review*, 39, 969–996.

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. (1996), "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

Chib, S. (1998), "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221–241.

Chib, S. and Greenberg, E. (2007), "Analysis of Additive Instrumental Variable Models," *Journal of Computational and Graphical Statistics*, 16, 86–114.

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

Chib, S., Greenberg, E., and Jeliazkov, I. (2009), "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Cogley, T. and Sargent, T. J. (2001), "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331–338.

Dahl, C. M. and Gonzalez-Rivera, G. (2003a), "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, Article 2.

Dahl, C. M. and Gonzalez-Rivera, G. (2003b), "Testing for Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields," *Journal of Econometrics*, 114, 141–164.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), "Automatic Bayesian Curve Fitting," *Journal of the Royal Statistical Society – Series B*, 60, 333–350.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., and Smith, A. F. M. (2002), *Bayesian Methods for Nonlinear Classification and Regression*, John Wiley & Sons, New York.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001), "Bayesian Curve-Fitting with Free-Knot Splines," *Biometrika*, 88, 1055–1071.

Dueker, M. (2005), "Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of U.S. Recessions," *Journal of Business & Economic Statistics*, 23, 96–104.

Fahrmeir, L. and Lang, S. (2001), "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors," *Journal of the Royal Statistical Society – Series C*, 50, 201–220.

Gelfand, A. E. (2000), "Discussion to "Bayesian Backfitting"," *Statistical Science*, 15, 217–218.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis*, Chapman & Hall, New York, 2 edn.

Greenberg, E. (2008), *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hamilton, J. D. (2001), "A Parametric Approach to Flexible Nonlinear Inference," *Econometrica*, 69, 537–573.

Hansen, B. E. (1992), "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

Härdle, W. and Tsybakov, A. (1997), "Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression," *Journal of Econometrics*, 81, 223–243.

Härdle, W., Tsybakov, A., and Yang, L. (1998), "Nonparametric Vector Autoregression," *Journal of Statistical Planning and Inference*, 68, 221–245.

Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall, New York.

Hastie, T. and Tibshirani, R. (2000), "Bayesian Backfitting," *Statistical Science*, 15, 196–223.

Holmes, C. C., Denison, D. G. T., and Mallick, B. K. (2002), "Accounting for Model Uncertainty in Seemingly Unrelated Regressions," *Journal of Computational and Graphical Statistics*, 11, 533–551.

Imai, K. and van Dyk, D. (2005), "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation," *Journal of Econometrics*, 124, 311–334.

Jeliazkov, I. (2011), "Specification and Inference in Nonparametric Additive Regression," *working paper, Department of Economics, University of California, Irvine*.

Jeliazkov, I. and Lee, E. H. (2010), "MCMC Perspectives on Simulated Likelihood Estimation," *Advances in Econometrics: Maximum Simulated Likelihood*, 26, 3–39.

Kim, C.-J. and Nelson, C. R. (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics*, 81, 608–616.

Kim, C.-J., Morley, J., and Piger, J. (2005), "Nonlinearity and the Permanent Effects of Recessions," *Journal of Applied Econometrics*, 20, 291–309.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons, New York.

Koop, G. and Korobilis, D. (2009), "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends in Econometrics*, 3, 267–358.

Koop, G. and Poirier, D. J. (2004), "Bayesian Variants of Some Classical Semiparametric Regression Techniques," *Journal of Econometrics*, 123, 259–282.

Koop, G., Poirier, D. J., and Tobias, J. (2005), "Bayesian Semiparametric Inference in Multiple Equation Models," *Journal of Applied Econometrics*, 20, 723–747.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003), "International Business Cycles: World, Region and Country Specific Factors," *American Economic Review*, 93, 1216–1239.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2008), "Understanding the Evolution of World Business Cycles," *Journal of International Economics*, 75, 110–130.

Lin, X. and Zhang, D. (1999), "Inference in Generalized Additive Mixed Models by Using Smoothing Splines," *Journal of the Royal Statistical Society – Series B*, 61, 381–400.

Lindley, D. V. (1971), *Bayesian Statistics: A Review*, SIAM, Philadelphia.

Meng, X.-L. and van Dyk, D. (1999), "Seeking Efficient Data Augmentation Schemes via Conditional and Marginal Augmentation," *Biometrika*, 86, 301–320.

Pesaran, M. H. and Potter, S. M. (1997), "A Floor and Ceiling Model of U.S. Output," *Journal of Economic Dynamics and Control*, 21, 661–695.

Poirier, D. J. (1973), "Piecewise Regression Using Cubic Spline," *Journal of the American Statistical Association*, 68, 515–524.

Poirier, D. J. (1998), "Revising Beliefs in Non-Identified Models," *Econometric Theory*, 14, 483–509.

Potter, S. M. (1995), "A Nonlinear Approach to US GNP," *Journal of Applied Econometrics*, 10, 109–125.

Primiceri, G. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–852.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press, Cambridge, UK.

Shiller, R. (1984), "Smoothness Priors and Nonlinear Regression," *Journal of the American Statistical Association*, 79, 609–615.

Shively, T. S., Kohn, R., and Wood, S. (1999), "Variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based Prior," *Journal of the American Statistical Association*, 94, 777–806.

Silverman, B. (1985), "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," *Journal of the Royal Statistical Society – Series B*, 47, 1–52.

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

Sims, C. A. and Zha, T. (2006), "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96, 54–81.

Smith, M. and Kohn, R. (2000), "Nonparametric Seemingly Unrelated Regression," *Journal of Econometrics*, 98, 257–281.

Stock, J. H. and Watson, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Stock, J. H. and Watson, M. W. (2003), "Has the Business Cycles Changed? Evidence and Explanations," in *Monetary Policy and Uncertainty: Adapting to a Changing Economy*, pp. 9–56, Federal Reserve Bank of Kansas City.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.

van Dyk, D. and Meng, X.-L. (2001), "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50.

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*, 40, 364–372.

Wasserman, L. (2006), *All of Nonparametric Statistics*, Springer, New York.

Whittaker, E. (1923), "On a New Method of Graduation," *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75.

Wood, S. and Kohn, R. (1998), "A Bayesian Approach to Robust Binary Nonparametric Regression," *Journal of the American Statistical Association*, 93, 203–213.

Wood, S., Kohn, R., Shively, T., and Jiang, W. (2002), "Model Selection in Spline Nonparametric Regression," *Journal of the Royal Statistical Society – Series B*, 64, 119–139.

Yang, L., Härdle, W., and Nielsen, J. (1999), "Nonparametric Autoregression with Multiplicative Volatility and Additive Mean," *Journal of Time Series Analsys*, 20, 579–604.