Journal of Econometrics 159 (2010) 134–150

Contents lists available at ScienceDirect

Journal of Econometrics

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

A flexible approach to parametric inference in nonlinear and time varying time series models☆

Gary Koop [a,*], Simon Potter [b]

[a] Department of Economics, University of Strathclyde, United Kingdom

[b] Federal Reserve Bank of New York, United States

ARTICLE INFO

Article history:
Received 2 August 2007
Received in revised form
18 November 2009
Accepted 11 May 2010
Available online 15 June 2010

JEL classification:
C11
C22
E17

Keywords:
Bayesian
Structural break
Threshold autoregressive
Regime switching
State space model

ABSTRACT

Many structural break and regime-switching models have been used with macroeconomic and financial data. In this paper, we develop an extremely flexible modeling approach which can accommodate virtually any of these specifications. We build on earlier work showing the relationship between flexible functional forms and random variation in parameters. Our contribution is based around the use of priors on the time variation that is developed from considering a hypothetical reordering of the data and distance between neighboring (reordered) observations. The range of priors produced in this way can accommodate a wide variety of nonlinear time series models, including those with regime-switching and structural breaks. By allowing the amount of random variation in parameters to depend on the distance between (reordered) observations, the parameters can evolve in a wide variety of ways, allowing for everything from models exhibiting abrupt change (e.g. threshold autoregressive models or standard structural break models) to those which allow for a gradual evolution of parameters (e.g. smooth transition autoregressive models or time varying parameter models). Bayesian econometric methods for inference are developed for estimating the distance function and types of hypothetical reordering. Conditional on a hypothetical reordering and distance function, a simple reordering of the actual data allows us to estimate our models with standard state space methods by a simple adjustment to the measurement equation. We use artificial data to show the advantages of our approach, before providing two empirical illustrations involving the modeling of real GDP growth.

1. Introduction

Many recent developments in empirical macroeconomics are based on statistical models which are nonlinear or exhibit structural breaks or time variation in parameters. For instance, Cogley and Sargent (2001, 2005), Boivin and Giannoni (2006) and Primiceri (2005) use structural break or time varying parameter models to examine whether monetary policy rules have changed over time. Other authors (e.g. Sims and Zha, 2006; Koop and Potter, 2006) develop regime-switching models to analyze similar issues relating to monetary policy. The potential empirical importance of departures from constant parameter linear models is undeniable. However, in practice, a problem arises since the set of possible models which exhibit time variation or regime-switching in coefficients is huge and the potential for data mining is commensurately large. These considerations have led to an interest in developing flexible parametric models[1] which nest common nonlinear time series and time varying parameter specifications (see, e.g. Hamilton, 2001, 2003; Lundbergh et al., 2003; Bec et al., 2008; Giordani et al., 2007). Our paper adds a new approach to flexible modeling to complement this existing literature but, we argue, embeds more flexibility than previous work.

The intuition underlying our modeling framework is based on two ideas: hypothetical data reordering and distance between (reordered) observations. The first can be motivated by comparing an autoregressive model for y_t with a structural break at time τ and a threshold autoregressive (TAR) model with threshold τ (i.e. the AR dynamics, when $y_{t-1} < \tau$, are different from those when $y_{t-1} \geq \tau$). Suppose we create a new variable y_t^* which is a simple reordering of y_t according to y_{t-1} (i.e. y_1^* is y_s where

☆ We would like to thank Paolo Giordani, Arnold Zellner, an associate editor, referees and seminar participants at the Rimini Centre for Economic Analysis and the University of Nottingham for numerous insightful and helpful comments. The views expressed in this paper are those of the authors and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System. Gary Koop is a Fellow of the Rimini Centre for Economic Analysis. Gary Koop acknowledges financial support from the Leverhulme Trust under Grant F/00 273/J.

* Corresponding author.
E-mail addresses: Gary.Koop@strath.ac.uk (G. Koop), Simon.Potter@ny.frb.org (S. Potter).

[1] Nonparametric approaches are also a promising avenue, although they are less popular in this literature since macroeconomic data sets tend not to be that large (see also the reasons outlined in the introduction to Hamilton, 2001).

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.05.002

y_{s-1} is the smallest value of lagged y, y_2^* has the second smallest value for lagged y, etc.). Then the structural break and TAR models are statistically exactly the same model, but one uses the original data y_t and the other uses the hypothetical reordered data y_t^*. In general, a wide range of nonlinear time series models with regime-switches can be re-interpreted as structural break models using hypothetical reordered data. An equivalent way of stating this is that many structural break models and regime-switching models can be viewed as differing in the hierarchical priors placed on the coefficients. But if the data for regime-switching models were to be reordered then the two priors on the coefficients are equivalent. Thus, the same algorithm can be used to estimate the structural break model and the regime-switching model if one reorders the data for the latter.

Hamilton (2001) develops a flexible approach to nonlinearity using random variation in parameters. There is a related literature that compares state space models of time variation in parameters to nonparametric and spline methods (see Harvey and Koopman, 2000). Hypothetical data reordering links these two literatures. We add a new element by introducing more flexibility in how the variation is modeled based on the distance between observations. Current approaches to random variation in parameters are very similar to nonparametric regression methods (see, e.g. Yatchew, 1998): when approximating $y = f(x)$, if two observations have similar values for x (i.e. the distance between their values for x is small) then both random variation in parameter methods and nonparametric regression methods imply that they have similar values for $f(x)$. We maintain this assumption locally but allow for substantial heterogeneity in time variation globally. For example, consider modeling changes in dynamics over the business cycle using an autoregressive model. In this case, we might want the autoregressive coefficients to change quickly (i.e. exhibit substantial variation) around business cycle turning points, but change more slowly during an expansion phase. Our approach allows for this. Or consider an example from a different literature: time series with seasonal patterns. Here the reordering of the data is no longer hypothetical as it relates to the well-defined concept of a particular seasons (i.e., quarters, months or weeks). It is assumed that seasons close to each other have similar dynamics (i.e., February is more like January than July).

In the next section, we provide more detail on why combining these two ideas–hypothetical data reordering and distance between (reordered) observations—results in a modeling framework that is flexible and nests virtually every popular parametric structural break and regime-switching model. We argue that an advantage of our approach is that, instead of assuming a particular model (e.g. a TAR or an AR model with a fixed number of breaks), it can allow the data to tell us which (if any) departure from linearity is appropriate. Thus, our approach is flexible. However, it is also computationally straightforward. We remain in a familiar class of models which are easy to understand and easy to handle econometrically. That is, conditional on a distance function and ordering, every model we consider is a state space model once the measurement equation is given a suitable interpretation. Since Bayesian methods for state space models are well developed, we can use such methods and only add a block to an existing posterior simulator which characterizes the distance function. In Section 3 of this paper, we describe such a posterior simulation algorithm for an empirically-relevant implementation of our approach. This applies our modeling framework to both conditional mean and volatility parameters. Section 4 contains empirical work. After illustrating our approach with artificial data we apply our techniques to the modeling of US GDP growth. We start with the univariate time series properties of GDP growth. We find overwhelming evidence in favor of time variation in the conditional variance with a form similar to stochastic volatility. We then add lags of oil price inflation as covariates with time varying coefficients. In this application, we find evidence in favor of a regime-switching model where oil price changes averaged over the past year trigger a switch between regimes (but with substantial variation within each regime).

2. A flexible parametric modeling framework

In this section, we outline the general features of our modeling framework in terms of a simple variant of our model. In the following section we introduce a more general model suitable for empirical research. In this intuitive section, we assume that the error variance is homoskedastic and do not discuss volatility issues. However, our general model described in the next section does apply our flexible parametric framework to volatility issues.

Consider a time varying parameter (TVP) model written in state space form with measurement equation given by:

$$y_t = \theta_t x_t + \sigma_\varepsilon \varepsilon_t, \tag{1}$$

for $t = 1, \ldots, T$ where x_t is a scalar (e.g. a lag of the dependent variable) and state equation given by:

$$\theta_t = \theta_{t-1} + \sigma_v v_t \tag{2}$$

and ε_t and v_t are i.i.d. $N(0, 1)$ (and independent of one another). This model is of interest in and of itself (e.g. Cogley and Sargent, 2001, use a VAR extension of it), nests some interesting models (e.g. $\sigma_v = 0$ is the linear model) and textbook methods for statistical inference (e.g. MCMC algorithms) are available. The relationship between state space models such as (1) and (2) and nonparametric kernel smoothing algorithms is well developed in the state space literature (e.g. Harvey, 1989; Harvey and Koopman, 2000). State space models such as (1) and (2) are a flexible and powerful tool for time series analysis. They are included in our approach as a special case.

At the outset, we should make clear that there are two ways of motivating our framework. These are equivalent motivations but use different terminology. One way, which can be thought of as the conventional Bayesian motivation, involves retaining the measurement equation throughout, but conceptualizing our approach as replacing the hierarchical prior defined by (2) with different hierarchical priors which define different nonlinear time series models. The second way, which adopts the language of the state space modeler and motivates our approach as retaining the form of the conventional state space model (and, thus, conventional econometric methods developed for state space models), but using hypothetically reordered observations. In this paper, we use both motivations, hoping to appeal to an audience of both Bayesians and state space modelers. But we stress that they are equivalent to one another. We begin with a Bayesian motivation and use notation where $y = (y_1, \ldots, y_T)'$, $\theta = (\theta_1', \ldots, \theta_T')'$ and $\lambda = (\sigma_\varepsilon, \sigma_v, \theta_0)'$ denotes the other model parameters.

To define a model we need the following distributions: $p(y|\theta, \lambda)$, $p(\theta|\lambda)$ and $p(\lambda)$. Our models are all identical in $p(y|\theta, \lambda)$ which is defined by (1) and differ only in $p(\theta|\lambda)$.[2] To provide some motivation for our extension of the model given by (1) and (2), it is useful to think in terms of $p(\theta|\lambda)$ which is the hierarchical prior defined by the state equation, (2). This is:

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{T-1} \\ \theta_T \end{bmatrix} \Bigg| \lambda \sim N\left(\begin{bmatrix} \theta_0 \\ \theta_0 \\ \vdots \\ \theta_0 \\ \theta_0 \end{bmatrix}, \sigma_v^2 \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & \ddots & 2 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & \ddots & T-1 & T-1 \\ 1 & 2 & \cdots & T-1 & T \end{bmatrix} \right).$$

[2] We also assume $p(\lambda)$ is identical in every model (i.e., different orderings and distance functions) for the same time series. Allowing for $p(\lambda)$ to differ across models for the same time series is an obvious extension of our approach.

One way of interpreting what we do in this paper is maintain the framework given by (1) and (2) (and, thus, still use standard MCMC algorithms) but investigate ways of parameterizing a general covariance matrix for the prior for the states. The most general prior covariance matrix would be:

$$var\begin{bmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_{T-1}\\ \theta_T\end{bmatrix}\Bigg|\lambda = \begin{bmatrix}\sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1T-1} & \sigma_{1T}\\ \sigma_{12} & \sigma_{22}^2 & \ddots & \sigma_{2T-1} & \sigma_{2T}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \sigma_{1T-1} & \sigma_{2T-1} & \ddots & \sigma_{T-1T-1}^2 & \sigma_{T-1T}\\ \sigma_{1T} & \sigma_{2T} & \cdots & \sigma_{T-1T} & \sigma_{TT}^2\end{bmatrix}.$$

Without further restrictions on this prior covariance matrix standard MCMC algorithms are not available and the model would be over-parameterized. Hence, we need some restrictions. A general way of expressing a wide class of restrictions is by introducing a sequence of realizations of random variables $z = (z_1, \ldots, z_T)'$ and replacing $p(\theta|\lambda)$ by $p(\theta|\lambda, z)$. A standard example of such as approach arises if we use stochastic volatility to model the scaling of the innovation in the state equation (see Stock and Watson, 2007). In this case if $\{z_t\}$ is the realized value for the log stochastic volatility sequence we would have:

$$var\left[\begin{array}{c|c}\theta_1\\ \theta_2\\ \vdots & \lambda, z\\ \theta_{T-1}\\ \theta_T\end{array}\right] = \sigma_v^2$$

$$\times\begin{bmatrix}\exp(z_1) & \exp(z_1) & \cdots & \exp(z_1) & \exp(z_1)\\ \exp(z_1) & \exp(z_1+z_2) & \ddots & \exp(z_1+z_2) & \exp(z_1+z_2)\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \exp(z_1) & \exp(z_1+z_2) & \ddots & \exp\left(\sum_{t=1}^{T-1} z_t\right) & \exp\left(\sum_{t=1}^{T-1} z_t\right)\\ \exp(z_1) & \exp(z_1+z_2) & \cdots & \exp\left(\sum_{t=1}^{T-1} z_t\right) & \exp\left(\sum_{t=1}^{T} z_t\right)\end{bmatrix},$$

and conditional on z standard algorithms can be used as this prior covariance matrix can be expressed as a state equation with the innovation scaled by $\exp(z_t/2)$. Now consider a sequence of stochastic volatility where $z_s = 0$ except at one time t. Then the prior variance matrix would be:

$$var\left[\begin{array}{c|c}\theta_1\\ \vdots\\ \theta_t & \lambda, z_t\\ \vdots\\ \theta_T\end{array}\right] = \sigma_v^2\begin{bmatrix}1 & \cdots & 1 & \cdots & 1\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ 1 & \cdots & \exp(z_t) & \cdots & \exp(z_t)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ 1 & \cdots & \exp(z_t) & \cdots & \exp(z_t)\end{bmatrix}.$$

This would define a single break model with the break at time t with the only non-zero innovation to the state equation at time t. To see this note that such an unconditional prior covariance matrix implies conditional prior variances of the form:

$$var(\theta_s|\theta_{s-1}) = 0,$$

for $s \neq t$, but $var(\theta_t|\theta_{t-1}) \neq 0$.

A general class of structures, of similar form to the above two examples, could be produced by constructing a function over the time index and using this to scale the innovation in the state equation. If we let $d(t, t-1) \geq 0$ be this scaling then we have a prior covariance matrix as given in Box I.

We call $d(t, t-1)$ the distance function and one of the features of our flexible parametric model is to extend basic state space models such as (1) and (2) to allow for this. The discussion above shows that, if $d(t, t-1)$ is equal to 1 for all time points we are back to the original TVP model, if $d(t, t-1)$ is zero except at one point we have the single break model. If we give $d(t, t-1)$ a flexible function form we can mimic the behavior of the stochastic volatility model or multiple break models. In the next section, we will offer further discussion on the flexibility that such an addition provides us with.

A second extension of our model relates to the hypothetical reordering of the data. Note that $p(\theta|\lambda, z)$ does not have to be expressed in the standard time ordering. For example, consider a non-decreasing sequence $\{z_s\}$ and associated distance function $d(z_z, z_{s-1})$; then a valid prior covariance matrix is as given in Box II.

At this stage, it is worth stressing that the previous material is intended to illustrate how many Bayesian models can be defined if we begin with (1) and use a multivariate Normal prior for $p(\theta|\lambda, z)$ but then consider different covariance matrices for this prior. So far this motivation has been purely Bayesian. But for the state space modeler, note that the preceding prior covariance matrix is equivalent to that produced by a state equation of the form:

$$\theta_s = \theta_{s-1} + \sigma_v d(z_s, z_{s-1})\, v_s, \tag{3}$$

where s indexes a reordering (in ascending order) of time according to the function of an *index variable* z_t (e.g. z_t could be an observed exogenous variable such as a lagged dependent variable). In the remainder of this paper, t will always be used as the natural time ordering (i.e. $t = 1, 2, \ldots, T$) and s will always denote a reordering according to an index variable. So, for instance, if $T = 3$ and $z_1 = 0.02, z_2 = 0.01$ and $z_3 = 0.03$, then $t = 1, 2, 3$ and $s = 2, 1, 3$. We will let the $T \times 1$ parameter γ define the ordering of the data. For instance, in the simple 3 period example when z_t is used to order the data we have $\gamma = (2, 1, 3)'$.

We also stress that this reordering notation applies to z so that z_s and z_{s-1} will be adjacent observations (e.g. if the index is lagged output growth, then z_{s-1} will have the next lowest value of lagged output growth to z_s). Different choices for γ will define different models. Furthermore, we treat $d = (d(z_2, z_1), d(z_3, z_2), \ldots, d(z_T, z_{T-1}))'$ as a vector of parameters (with hierarchical priors) and, thus, our methods allow for their data-based estimation.

Thus, from a Bayesian point of view, we can see that different priors for controlling parameter variation lead to a huge range of models and that these priors can be conceptualized as arising from hypothetical reorderings of the data. From the point of view of the state space modeler, we can see that different reorderings of the data and different distance functions can be used to nest a wide range of common specifications. The following discussion will make clear these points. In terms of computation, note that conditional on the ordering and distance function we are back to the standard state space framework if we combine state equation (3) with measurement equation occurring according to the time ordering s:

$$y_s = \theta_s x_s + \sigma_\varepsilon \varepsilon_s. \tag{4}$$

Thus, standard methods of posterior simulation can be used given the time ordering and distance function. Further, using Bayesian modeling averaging one can combine the results from models using different orderings and distance functions.

In order to further explain and motivate our models, we first consider the role played by the distance function, then the role played by data reordering, before finally combining both aspects together.

2.1. The role of the distance function

Consider first, the sorts of models which result if the data are in standard time series order. In terms of our definitions, we have

$$var\left[\begin{array}{c}\theta_1\\ \theta_2\\ \vdots\\ \theta_{T-1}\\ \theta_T\end{array}\middle|\ \lambda, \{z_t = t\}\right] = \sigma_v^2\begin{bmatrix} d^2(1,0) & d^2(1,0) & \cdots & d^2(1,0) & d^2(1,0) \\ d^2(1,0) & d^2(1,0)+d^2(2,1) & \ddots & d^2(1,0)+d^2(2,1) & d^2(1,0)+d^2(2,1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d^2(1,0) & d^2(1,0)+d^2(2,1) & \ddots & \sum_{t=1}^{T-1} d^2(t,t-1) & \sum_{t=1}^{T-1} d^2(t,t-1) \\ d^2(1,0) & d^2(1,0)+d^2(2,1) & \cdots & \sum_{t=1}^{T-1} d^2(t,t-1) & \sum_{t=1}^{T} d^2(t,t-1) \end{bmatrix}$$

Box I.

$$var\left[\begin{array}{c}\theta_1\\ \theta_2\\ \vdots\\ \theta_{S-1}\\ \theta_S\end{array}\middle|\ \lambda, \underline{z}, \{z_s\}\right] = \sigma_v^2\begin{bmatrix} d^2(z_1,z_0) & d^2(z_1,z_0) & \cdots & d^2(z_1,z_0) & d^2(z_1,z_0) \\ d^2(z_1,z_0) & d^2(z_1,\underline{z})+d^2(z_2,z_1) & \ddots & d^2(z_1,\underline{z})+d^2(z_2,z_1) & d^2(z_1,\underline{z})+d^2(z_2,z_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d^2 z_1,\underline{z}_0 & d^2(z_1,\underline{z})+d^2(z_2,z_1) & \ddots & \sum_{s=1}^{T-1} d^2(z_s,z_{s-1}) & \sum_{s=1}^{T-1} d^2(z_s,z_{s-1}) \\ d^2(z_1,z_0) & d^2(z_1,\underline{z})+d^2(z_2,z_1) & \cdots & \sum_{s=1}^{T-1} d^2(z_s,z_{s-1}) & \sum_{s=1}^{T} d^2(z_s,z_{s-1}) \end{bmatrix}$$

Box II.

$\gamma = (1, \ldots, T)'$, $z_t = t$ for $t = 1, \ldots, T$ (i.e. the index variable is already in ascending order and, thus, no reordering of the data is required to define an s). Thus, the model is:

$$y_t = \theta_t y_{t-1} + \sigma_\varepsilon \varepsilon_t, \tag{5}$$

where

$$\theta_t = \theta_{t-1} + \sigma_v d(t, t-1)\, v_t. \tag{6}$$

Three cases are immediately clear. First, a standard linear AR model is obtained if $d(t, t-1) = 0$ for all t. Second, a TVP model of the sort used by Cogley and Sargent (2001) or Koop and Potter (2001) is obtained if $d(t, t-1) = 1$ for all t. Thirdly, structural break models also fit in this framework. For instance, if we let $d(t, t-1) = 1$ if $t = \tau$ and $d(t, t-1) = 0$ otherwise, then $\theta_1 = \cdots = \theta_{\tau-1}$ and $\theta_\tau = \cdots = \theta_T$ and we have a model with a single structural break at τ. By treating τ as an unknown parameter, the breakpoint can occur at an unknown point in time. A second breakpoint of unknown timing can be obtained by defining two breakpoints as τ_1 and τ_2 (where $\tau_2 > \tau_1$) and letting $d(t, t-1) = 1$ if $t = \tau_1$ and $d(t, t-1) = \sigma_2$ if $t = \tau_2$ and $d(t, t-1) = 0$ for all other values of t. Models with more than two structural breaks can be obtained by extending this distance function definition in the obvious manner.

In addition to these special cases, by choosing flexible functional forms for the distance function, we can obtain a range of values for the innovation variance scaling which allow for smoothing of various sorts. Such a model would be "nonparametric in spirit" and, hence, we occasionally use this terminology below (although we stress that our model is always a parametric one). In our approach, it is possible to select distance functions which are analogous to kernels used in nonparametric smoothing. Related to this, it is worth noting that, in the context of state space modeling with irregularly spaced time series data, Harvey (1989) develops filtering and smoothing methods and Harvey and Koopman (2000) discusses their relationship to nonparametric regression methods (and we further elaborate on these below). With irregularly space data, the distance between observations matters in Harvey's derivations and our distance function plays an analogous role. See also Koop and Poirier (2004) for related Bayesian discussion.

Traditionally, the structural break literature has focussed on two extremes. One extreme assumes there a small number of breaks (and changes in coefficients are large and possibly heterogeneous), the other assumes that breaks occur in every time period (but changes in coefficients are small and homogeneous). Recently, through work such as Giordani and Kohn (2008) and Koop and Potter (2007), there is a growing interest in structural break models lying between these extremes. Our distance function allows for this. As shown above, it is even possible in our approach to assume the innovation in the state equation has stochastic volatility (see Stock and Watson, 2007). The point we are making here is that, even without reordering the data, by suitably re-defining the distance function we can cover all of these possibilities in a single modeling framework. Furthermore, instead of making a single choice of models within this huge set, we can choose a parametric form for $d(t, t-1)$ and let the data tell us its form.

In our "Further Discussion" section below (subsequent to our discussion of the role of hypothetical data reordering), we provide some theoretical derivations which provide additional insight to the role of the distance function and its relationship to nonparametric regression.

2.2. The role of hypothetical data reordering

By allowing for a hypothetical reordering of the data, we can accommodate various nonlinear time series models within a standard state space algorithm. It has long been recognized that many common nonlinear time series models are equivalent in a statistical estimation sense to models with structural breaks if the data is suitably reordered (e.g. Tsay, 1989). At a high level of generality, the idea that, by imagining reordering the data, we can define different models can be explained as follows.[3] Let $y^t = (y_1, \ldots, y_t)'$ denote the data through time t, $z^t = (z_1, \ldots, z_T)'$ be

[3] Obviously in some cases it is nonsensical to imagine data reorderings. For example, consider a time series with varies in a nonlinear manner around a deterministic time trend. Any imagined data reorderings for different nonlinear models would require the data to be detrended first.

an exogenous index variable,[4] $x^t = (x_1, \ldots, x_T)'$ be the exogenous variables in the measurement equation[5] and $\theta^t = (\theta_1, \ldots, \theta_t)'$. Then, extending our previous notation and allowing for the role of our index variables, our models all require specification of $p\left(y^T|\theta^T, z^T, x^T, \lambda\right)$, $p\left(\theta^T|z^T, x^T, \lambda\right)$ and $p(\lambda)$ where λ includes all other parameters in the model, including any which enter the distance function. Conventional state space models assume that $p\left(y^T|\theta^T, z^T, x^T, \lambda\right)$ can be decomposed as

$$p\left(y^T|\theta^T, z^T, x^T\right) = \prod_{t=1}^{T} p\left(y_t|\theta_t, z_t, x_t, \lambda\right).$$

If we also assume $p\left(\theta^T|z^T, x^T, \lambda\right) = p\left(\theta^T|z^T, \lambda\right)$ is defined by state equation (2), then standard algorithms for posterior simulation (e.g. involving the Kalman filter) exist for such models. But the right-hand side of this equation is a product and, thus, terms can be reordered in any manner without affecting the result. Hence, if s indexes a reordering of the data as defined below Eq. (3), then the preceding equation will be exactly the same as[6]:

$$p\left(y^T|\theta^T, z^T, x^T\right) = \prod_{s=1}^{T} p\left(y_s|x_s, \theta_s, z_s, \lambda\right).$$

Assume $p\left(\theta^T|z^T, \lambda\right)$ is defined by state equation $\theta_s = \theta_{s-1} + v_s$. It follows immediately that if we reorder the data and then use standard methods of posterior simulation for state space models using this reordered data, we will be carrying out valid posterior inference for the model defined by (1) with prior given by $\theta_s = \theta_{s-1} + v_s$. This is uncontroversial. The only questions are whether the priors (and, thus, models) motivated through such hypothetical data reorderings are empirically-interesting ones and how they relate to other nonlinear time series models. In the following paragraphs we discuss these questions.

To make things concrete, return to the general specification given in (3) and (4) and suppose y_t is real GDP growth and $x_t = z_t = y_{t-1}$. Then γ orders the data based on last period's GDP growth. If we define the distance function as $d(z_s, z_{s-1}) = 1$ if $z_{s-1} < \tau$ and $z_s \geq \tau$, and $d(z_s, z_{s-1}) = 0$ otherwise, then we obtain a two-regime TAR model (e.g. Potter, 1995):

$$y_t = \theta_1 y_{t-1} + \sigma_\varepsilon \varepsilon_t \quad \text{if } y_{t-1} < \tau$$
$$y_t = \theta_2 y_{t-1} + \sigma_\varepsilon \varepsilon_t \quad \text{if } y_{t-1} \geq \tau.$$

Multiple regime TAR models involve the obvious extension of this distance function definition.

An important issue with threshold models is the choice of the index variable, z. In the standard implementation of the TAR model, the index variable is a lag of the dependent variable. By defining z_t to be other functions of lagged dependent variables (or other exogenous variables), we can get a wide range of TAR models as, e.g., in Koop and Potter (1999). This is an avenue we pursue in our empirical work.

These examples give only a flavor of the wide variety of behaviors allowed by this flexible parametric model. Indeed, if γ were to be totally unrestricted and we consider a flexible form for the distance functions then this class of models would be so flexible as to be virtually equivalent to a nonparametric model. However, if γ is left completely unrestricted, then the number of configurations it could take is $T!$ (i.e. there are $T!$ ways of reordering the data). For most macroeconomic data sets, $T!$ is simply too large to allow for exhaustive consideration of all possible hypothetical reorderings of the data.[7] Accordingly, in the empirical work done in this paper we restrict the set of allowable configurations for γ.

2.3. Cases involving latent variables

In the previous sections we assumed that z_t is an observed variable such as a lagged dependent variable or an exogenous variable (and our empirical work will always make this assumption). However, it is worth digressing briefly to emphasize that our approach can be used when z_t is a latent variable and some very interesting models result. Furthermore, we show how adding new latent variables can extend our model to be similar to other popular models.

Consider first some structural break models. The structural break framework described above involves a fixed (known) number of structural breaks. That is, we define a distance function for each possible number of breaks (e.g. one distance function defines a one break model, another defines a two break model, etc.). If the number of breaks is unknown, we could simply do Bayesian model averaging over models with differing numbers of breaks. However, for reasons discussed in Koop and Potter (2007), it can be desirable to work with a model which does not impose a fixed number of breakpoints. This can be done by adopting a hierarchical prior for $d(t, t-1)$ of a specific sort. This brings us into the family of models developed in McCulloch and Tsay (1993), Gerlach et al. (2000) and Giordani and Kohn (2008). As a simple example, we can introduce a sequence of latent variables, $\{s_t\}$, where $s_t \in \{0, 1\}$, and let $d(t, t-1; s_t) \equiv s_t d(t, t-1)$ for $t = 1, \ldots, T$ be unknown parameters and use a hierarchical Bernoulli prior distribution for s_t:

$$p(s_t = 1) = p,$$
$$p(s_t = 0) = 1 - p$$

where p is an unknown parameter. This is the model of McCulloch and Tsay (1993) if $d(t, t-1) = 1$. Note that this allows for a break to occur in every period with probability p and, thus, the number of breaks can be estimated in the data. This model can be extended to allow for p to change if a break occurs, p to depend on exogenous variables, etc. Furthermore, in more general models s_t can be a vector (e.g. it can have two components, one controlling breaks in coefficients and the other in error variances).[8] As long as the hierarchical prior for s_t has a Markov structure (with independence being a special case of this), the efficient algorithms of Gerlach et al. (2000) and Giordani and Kohn (2008) can be used to obtain posterior draws of the sequence $\{s_t\}$.

Structural break models can also be expressed directly in terms of latent variables z_t which can take on a range of integer values denoting the different regimes in the data. So, for instance, when $T = 5$ then $z^T = (1, 1, 1, 2, 2)'$ denotes a structural break after $t = 3$. The structural break model of Chib (1998) involves latent

[4] As discussed in the next section, they can also be realizations of latent variables produced by a posterior simulator.

[5] These can be lagged dependent variables.

[6] To avoid confusion, note that it is often the case that x_t contains lagged dependent variables. For instance, in the AR(1) case y_t depends on y_{t-1}. When we are using reordered data then y_s depends on x_s. This still implies y_s depends on its lag (in natural time ordering). It does not imply that y_s depends on its next lowest neighbor in the ordering.

[7] The issues raised by having $T!$ configurations of γ are essentially the same as those raised in a Bayesian model averaging or selection exercise involving regressions where the number of potential explanatory variables, K, is large and the number of models is 2^K. In such cases, simulation algorithms over model space (see, e.g., Hoeting et al., 1999 or Chipman et al., 2001) can be used which do not require exhaustive evaluation of every model. In some cases, these might be useful for the present class of models. However, for values of T greater than approximately 20 the computational demands of such simulation algorithms are currently high.

[8] An extension we do not consider is to a structural break model where only some of the conditional mean parameters change as in Levin and Piger (2008). This could be captured in our approach by modifying the measurement equation to allow for some explanatory variables to have constant coefficients.

variables of this form with restrictions on the matrix of Markov transition probabilities:

$$p(z_t = j|z_{t-1} = j-1) = p_{j-1},$$
$$p(z_t < j-1|z_{t-1} = j-1) = 0,$$
$$p(z_t > j|z_{t-1} = j-1) = 0.$$

Note that the latent variables, z^T, defined in this way are always in non-descending order and $d(z_t, z_{t-1})$ is automatically either zero (when no break occurs) or one (when a break occurs). Heterogeneous break sizes can be obtained in a similar fashion as described above for the simple structural break model (see the discussion after Eq. (6)).

The previous discussion suggests that we can also obtain a Markov switching model by using an unrestricted matrix of Markov transition probabilities. That is, if z_t can take on the integer values $\{1, \ldots, M\}$ denoting M regimes in the data and the matrix of transition probabilities has the usual form based on:

$$p(z_t = j|z_{t-1} = i) = p_{i,j}$$

then we obtain the Markov switching model. In terms of an MCMC algorithm, treatment of this case proceeds as follows. Conditional on a draw of z^T, we imagine reordering all the data according to this variable, and then take a draw of the states and other model parameters as described previously. Note that, once the data is imagined "reordered" we have $d(z_s, z_{s-1})$ being either zero (when no regime-switch occurs) or one (when a regime-switch occurs). Furthermore, the argument made above proves that this is a valid way of drawing all model parameters (conditional on z^T). The MCMC algorithm is completed by drawing from z^T conditional on the model parameters. But (if the data and z^T is placed back in natural time ordering) then standard methods for drawing from Markov switching models can be used (see, e.g. Chib, 1996).

2.4. Further discussion

In this section, we investigate further the types of nonlinear time series models we can obtain by combining hypothetical reorderings with distance functions as well as providing some theoretical properties of our approach. Note that ours is a hierarchical modeling approach that does not allow us to directly give a simple analytical representation of the properties of our nonlinear model (i.e. we cannot integrate out λ analytically).[9] However, if we focus on the lowest level of the hierarchy (i.e. provide results conditional on λ), then we can provide a number of useful illustrations of the flexibility of our approach in terms of what it implies for the conditional expectation and variance functions. But we do stress that our approach involves averaging over a range of possible conditional expectation and variance functions (i.e. integrating out λ).

The lowest level of our hierarchy involves $p\left(\theta^T|z^T, \lambda\right)$ which provides a prior for any period (say θ_s) conditional on a choice of distance function, ordering and all unknown parameters. Define the information set I_{-s} to include all information other than the values of θ_s and y_s. Then we have:

$$E[y_s|I_{-s}] = E[\theta_s|I_{-s}]x_s.$$

What is our model for conditional expectation function $E[\theta_s|I_{-s}]x_s$? In our approach, we assume that it is related to the value of a single index z_s. Thus unlike some very general nonparametric methods or the random field approach of Hamilton (2001), we limit ourselves to indices of one dimension. But such single index restrictions are often used to reduce the curse of dimensionality in many nonparametric approaches (and it is not be too difficult to extend our approach to more than one dimension). And it is important to stress that a strength of our approach is that we examine a wide range of choices for the single index.

Now let us consider adjacent values of the single index around z_s. Call these $\underline{z} < z_s$ and $\overline{z} > z_s$ and, corresponding to these, $\underline{\theta}$ and $\overline{\theta}$. According to our model, we have the following two relationships:

$$\theta_s = \underline{\theta} + \sigma_v d(z_s, \underline{z})v_t,$$
$$\overline{\theta} = \theta_s + \sigma_v d(\overline{z}, z_s)\overline{v},$$

where v_s and $\overline{v}$ are independent standard Normal random variables. Thus, our model says that θ_s has a Normal distribution with mean

$$\frac{\frac{\underline{\theta}}{d(z_s,\underline{z})^2} + \frac{\overline{\theta}}{d(\overline{z},z_s)^2}}{\frac{1}{d(z_s,\underline{z})^2} + \frac{1}{d(\overline{z},z_s)^2}}$$

and variance

$$\frac{\sigma_v^2}{\frac{1}{d(z_s,\underline{z})^2} + \frac{1}{d(\overline{z},z_s)^2}}.$$

From these equations, it can be seen how θ_s is a weighted average of adjacent observations. These weights can be examined for special cases to see how our approach can nest popular regime-switching and structural break models.

If d is only non-zero at only one point (say z^*) we have (depending on the definition of z) either a standard single structural break model (if z_s is simply time) or a two-regime threshold model/Markov switching model (if z is an appropriately defined time series variable, such as a lagged dependent variable). To be precise, it can be seen that:

1. if $z_s, \underline{z}, \overline{z} < z^*$ or $z_s, \underline{z}, \overline{z} > z^* \underline{\theta} = \theta_s = \overline{\theta}$.
2. if $\underline{z} < z_s < z^* \leq \overline{z}, \theta_s = \underline{\theta}$.
3. if $\underline{z} \leq z^* < z_s < \overline{z}, \theta_s = \overline{\theta}$.

A second case of interest is when d is positive and constant across $[\underline{z}, \overline{z}]$. In this case, the mean of θ_s is the simple average of $\underline{\theta}, \overline{\theta}$. This is very similar to the model of Hamilton (2001) restricted to one dimension. However, unlike Hamilton, our methods, by the use of a single index, directly produce an estimate of a model with a certain class (e.g. TAR, TVP, etc.) and combine the results over different classes of model (i.e. we will average over different choices of the single index z). Hamilton (2001) uses a frequentist testing approach to assess the presence of nonlinearity. This leads to all the usual sequential testing problems if there is more than one nonlinear specification one is interested in testing. Note that Hamilton's approach requires choosing the variables that might produce the nonlinearity. As is common in Bayesian work, we can avoid testing and instead do model averaging. The empirical application we give below (involving the oil price) shows the advantages of our approach in this regard. In it, there are many possible index variables that could be associated with nonlinearity. Doing many sequential tests, involving each of these variables, could run into pre-testing problems.

With a constant distance function, we have the following cases:

1. If z_s is simply time then we have the standard TVP model.
2. If z_s is another time series variable then we have a nonlinear time series model. The model could be Markov switching or threshold type, but with the assumption of a constant conditional expectation function in each regime is relaxed.

If d is positive and varies over $[\underline{z}, \overline{z}]$ then the mean of θ_s depends, via the distance function, on the weights on $\underline{\theta}, \overline{\theta}$. But we stress

[9] We define nonlinear as any departure from a linear model with time-invariant parameters.

that, unlike some nonparametric approaches, these weights are estimated. That is, so far we have focussed on the lowest level of the hierarchical model (i.e. the formula above are conditional on the distance function and other parameters in λ). Higher up the hierarchy, parameters in the distance function are estimated (or integrated out). Thus we can capture cases where the data suggests the distance function is close to one of the restricted forms described common in the nonlinear time series literature. We have the following cases:

1. If z is time we have new type of time varying parameter model with heterogeneous innovations. For example, d could include information on position of likely breaks. This model complements some new approaches to using heterogeneous innovations in the transition equation of TVP models (see Cogley and Sargent, 2007; Stock and Watson, 2007; Koop and Potter, 2007).
2. If d is not constant and z is a lag of the time series then we have a new type of nonlinear time series model. The details of the shape of the distance function will determine the type of nonlinearity.

Our model is flexible at this lowest level of hierarchy in the sense of the frequentist work of Hamilton (2001) on random fields because it can adapt to the observed data at the time point s by combining this prior view of θ_s with the observed data. The extent of updating in Bayes rule depends on precision of the prior for θ_s versus $1/\sigma_v$. And the full model (including all levels of the hierarchy) is more flexible still.

We stress that we allow the conditional variance function to have a flexible form and allow it to be separately identified from the conditional mean. Such generality is not available in the approach of Hamilton (2001) or other related nonparametric work. We can achieve this generality by our use of a single index restriction to limit dimensionality (although, of course, we can use a different single index for the conditional mean and conditional variance).

Finally, it is worth making clear the relationship between our methods and methods of nonparametric signal extraction. Harvey and Koopman (2000), Section 4, develop the links between state space methods, nonparametric methods and splines. The focus there is on the local linear trend model:

$$y_t = \mu_t + \sigma_\varepsilon \varepsilon_{1t},$$
$$\mu_{t+1} = \mu_t + \beta_t + \alpha\sigma_\varepsilon \varepsilon_{2t},$$
$$\beta_{t+1} = \beta_t + \gamma\sigma_\varepsilon \varepsilon_{3t}$$

where ε_{jt} for $j = 1, 2, 3$ are independent white noise with mean zero and variance one. When no hypothetical reordering of the data is used, our model is in some ways more flexible and in one way more restrictive than the local linear trend model. With regards to the latter point, given our focus on TVP regression models which typically involve random walk evolution of coefficients, we do not add the third state equation and so implicitly have $\gamma = 0$ (although it would be trivial for us to add this to our models). For the case where we have no explanatory variables in our measurement equation, but only an intercept, the discussion of Harvey and Koopman (2000) is of direct applicability (with $\gamma = 0$). As one example of the type of result they give is that, in the case where $\gamma = 0, \alpha^2$ plays a similar role to a kernel bandwidth. Such results will hold in our model when the no hypothetical reordering of the data is entertained and can be used to show the role played by our distance function. As another example, they discuss the equivalence[10] of the local linear trend model and the cubic spline. There is no exact equivalence with our methods here since we have $\gamma = 0$, but we could easily extend our model allow for such cubic spline behavior. These relationships (and many more) are further developed in, among many others, Durbin and Koopman (2001), Silverman (1985) and Wecker and Ansley (1983).

The preceding paragraph establishes that, for a simplified version of our model (i.e. with only an intercept in the measurement equation) and the data in natural time ordering, we can draw on a large literature linking state space models and nonparametric signal extraction methods to derive further properties of our approach. This material is also useful in understanding the properties of our approach when the data is not in natural time ordering but instead is ordered according to z_t. That is, the conventional nonparametric signal extraction approach can be interpreted as assuming a model of the form $y_t = f(t) + \varepsilon_t$ where $f(t)$ is the unknown trend that the researcher is seeking to estimate. Papers such as Harvey and Koopman (2000) discuss how state space methods estimate $f(t)$ and the resulting properties of the estimator. When the data is ordered according to z_t (and z_t are equally spaced), the estimator and its properties will take exactly the same form except as relating to the model $y_t = f(z_t) + \varepsilon_t$. For instance, results relating to how the local level model produces an estimated trend which smooths observations which are neighbors in time translate directly to results on how it smooths observations which are neighbors in terms of z_t. To allow for the fact that z_t are typically not spaced equally, our approach allows the distance between neighboring observations of z_t to enter the state equation error variance. By simply using this specification for the state equation error variance (or the implied signal to noise ratio) in the relevant formulae drawn from the literature (e.g. for how neighboring observations are weighted in a signal extraction exercise), the reader can easily derive the properties of our models as seen from the viewpoint of the signal extraction literature.

Table 1 provides an incomplete summary of the relationship between our modeling framework and the existing literature. Given our time series focus, we write this table assuming that the explanatory variables are lags of the dependent variable (i.e. x_t contains lags of the dependent variable). However, we stress that x_t could be any exogenous explanatory variables and, hence, our general framework holds for other regression-type models with regime-switches or structural breaks.

2.5. Specifying the distance function

The modeling framework so far holds for any distance function, although we have given a few specific examples that may be of empirical importance. In this section, we propose a particular implementation which should be flexible enough to let the data speak, but also be capable of accommodating the types of behaviors commonly observed with macroeconomic data. A convenient way of choosing the distance function is to think of it as being derived from a cumulative distribution function (CDF) on z_s. Then we have

$$d(z_s, z_{s-1}) \propto F(z_s) - F(z_{s-1}), \tag{7}$$

for some CDF $F(\cdot)$. If $F(\cdot)$ is based on a Uniform distribution (over the interval $[z_1, z_T]$) then the distance function reduces to $d(z_s, z_{s-1}) = z_s - z_{s-1}$. In this paper, we focus on the Normal distribution (with mean and variance estimated from the data using Bayesian methods). Note that the Normal is quite flexible when used in this context. The Normal can (by choosing a very large variance) approximate closely the $d(z_s, z_{s-1}) = z_s - z_{s-1}$ distance function. But it also can (by setting the mean to τ and choosing a very small variance) approximate closely the TAR distance function. Choosing the mean of the Normal in this case is, thus, analogous to estimating the threshold parameter in a

[10] Formally, this result requires the model to allow for ε_{2t} and ε_{3t} to be correlated.

Table 1
Links between our framework and popular nonlinear time series models.

Model	Distance function	Index variable
AR(p)	0	$z_t = t$
TVP	1	$z_t = t$
Structural Break 1 Break	$=1$ at time τ $=0$ otherwise	$z_t = t$
Structural Break K Breaks	$=1$ at $\tau_1, \ldots, \tau_K$ $=0$ otherwise	$z_t = t$
Structural Break Unknown # Breaks	$=1$ with prob p $=0$ otherwise	$z_t = t$
Chib (1998) Structural K Breaks Model	$=1$ with restricted Markov transition probs. $=0$ otherwise	$z_t = t$
Various nonparametric TVP models	Smooth function (e.g. kernel)	$z_t = t$
Standard TAR	$=1$ if $z_{s-1} < \tau$ and $z_s \geq \tau$ $=0$ otherwise	$z_t = y_{t-d}$
Other TARs	$=1$ if $z_{s-1} < \tau$ and $z_s \geq \tau$ $=0$ otherwise	z_t exogenous var. or functions of lags
Multiple Regime TARs	$=1$ if $z_{s-1} < \tau_1$ and $z_s \geq \tau_1$ $=1$ if $z_{s-1} < \tau_2$ and $z_s \geq \tau_2$ etc.	z_t exogenous var. or functions of lags
STAR[a]	Smooth function	$z_t = y_{t-d}$
Multiple Regime STAR	Smooth function with multiple modes	$z_t = y_{t-d}$
Markov switching model	$=1$ with restricted Markov transition probs. $=0$ otherwise	$z_t = t$
Various nonparametric time series models	Smooth function (e.g. kernel)	z_t exogenous var. or functions of lags

[a] This relationship is approximate and is illustrated in the artificial data section.

TAR model.[11] Intermediate values of the variance of the Normal would allow for a smooth change in dynamics around a threshold determined by the mean of the Normal (e.g. this would share some similarities with a smooth transition autoregressive or STAR model). The advantage of our approach is that the precise shape of the distance function would be estimated from the data and not imposed at the outset by choosing to estimate, e.g., a TAR or STAR model.

3. Bayesian inference

The previous discussion focussed on a simple state space model with homoskedastic errors. In our empirical work we begin with a TVP model of the form:

$$y_t = X_t'\theta_t + \varepsilon_t, \tag{8}$$

where y_t is a scalar and X_t is a vector with k elements (e.g. in our univariate application to real GDP growth, X_t includes a constant plus p lags of y_t). The model's coefficients evolve according to:

$$\theta_t = \theta_{t-1} + v_t \tag{9}$$

where $v_t \sim N(0, Q)$. The error in the measurement equation is assumed to exhibit stochastic volatility. That is,

$$\varepsilon_t = \xi_t \exp\left(\frac{1}{2}\alpha_t\right) \tag{10}$$

where $\xi_t \sim N(0, 1)$,

$$\alpha_t = \alpha_{t-1} + \eta_t \tag{11}$$

and $\eta_t \sim N\left(0, \sigma_\eta^2\right)$. The errors, ξ_t, v_t and η_t, are independent at all leads and lags and are independent of one another. Bayesian inference in this state space model can be done in a straightforward fashion using standard results. In our empirical work, we use the methods of Durbin and Koopman (2002) to draw from the posterior of $\theta = \left(\theta_0', \ldots, \theta_T'\right)'$ (conditional on $\alpha = (\alpha_0, \ldots, \alpha_T)'$ and Q).[12] The method of Kim et al. (1998) is used to draw from the posterior of α (conditional on θ and σ_η^2). To be precise, we use what they call their mixture sampler (see Section 3.3 of their paper). Kim et al. (1998) provide convincing evidence that this algorithm, since it draws directly from the posterior conditional for α, is much more efficient than single move samplers.

Note that, conditional on θ and α, the state equations reduce to simplified variants of linear regression models. Thus, conditional on draws of θ, the posterior for Q^{-1} takes the usual Wishart form (see, e.g. Koop, 2003, pages 140–141) and conditional on α the posterior of σ_η^{-2} takes the usual Gamma form (see, e.g. Koop, 2003, pages 61–62).

In summary, standard methods can be used to set up an MCMC algorithm which sequentially draws from $p(\theta|Data, \alpha, Q, \sigma_\eta^2)$, $p(\alpha|Data, \theta, Q, \sigma_\eta^2)$, $p(Q^{-1}|Data, \theta, \alpha, \sigma_\eta^2)$ and $p(\sigma_\eta^{-2}|Data, \theta, \alpha, Q)$.[13] The results of Fernandez et al. (1997) imply that proper

[11] Multiple regime models can be handled through mixtures of Normals. For example, an M regime TAR would require a mixture of $M + 1$ Normals with the mixture weights also being estimated parameters. In this paper, we do not investigate this extension although it is conceptually and computationally straightforward.

[12] Note that we use the algorithm on Durbin and Koopman (2002) to supply draws of $\theta_1, \ldots, \theta_T$. The initial condition, θ_0, can be treated as a regression effect and drawn (conditional on $\theta_1, \ldots, \theta_T$) using standard results for the Normal linear regression model.

[13] Here (and throughout this paper) we have written these out as the full posterior conditionals. However, some of them do not depend on all the conditioning

priors for either of the error variances or the initial conditions in state equations are required in order to obtain proper posteriors.[14] In our empirical work, we use weakly informative priors.

The previous discussion described an MCMC algorithm for a TVP model with stochastic volatility. However, we want to extend this model to allow for different hypothetical orderings and distance functions. Thus, analogously to (3) and (4), we define a variant of (8) and (9) which allows for reorderings of the data according to an index variable, z_t, and allows for the error variance in the transition equation to depend on the distance between observations (ordered according to the index variable). That is, our model is

$$y_s = X_s'\theta_s + \varepsilon_s, \tag{12}$$

where

$$\theta_s = \theta_{s-1} + \upsilon_s \tag{13}$$

and $\upsilon_s \sim N\left[0, \sqrt{d(z_s, z_{s-1})}Q\right]$. The error in the measurement equation has the form,

$$\varepsilon_s = \xi_s \exp\left(\frac{1}{2}\alpha_s\right) \tag{14}$$

where

$$\alpha_s = \alpha_{s-1} + \eta_s \tag{15}$$

and $\eta_s \sim N\left[0, \sqrt{d(z_s, z_{s-1})}\sigma_\eta^2\right]$. Additional assumptions about the errors are as described above (see Eqs. (8)–(11) and surrounding discussion). Other definitions are as after (4). Most importantly, remember that s indexes a hypothetical reordering of time according to the index variable, z_t, when $z_1, \ldots, z_T$ are placed in ascending order. We use s subscripts to denote the reordering so that z_s and z_{s-1} are adjacent observations. We introduce a parameter vector γ to denote how the index variable orders the data, the τth element of γ is the rank of z_τ in the ascending list of ordered data. $d(z_s, z_{s-1})$ is a (non-negative) distance function measuring the distance between z_s and z_{s-1}.

In some applications, it may be desirable to use a different hypothetical ordering and distance function for the conditional mean coefficients than the conditional variances. This is conceptually straightforward and is implemented in our second empirical example below. Note that, if we consider a large number of different hypothetical pairs of orderings for both the conditional mean and conditional variance, the computational demands would increase greatly. As we shall explain in the empirical section, in many cases involving macroeconomic data one could focus on the case where the modeling of conditional variance does not require any hypothetical reorderings.

We use a distance function based on Normal CDFs (see Eq. (7)). If we let $\Phi\left(z_s; \mu, \sigma_d^2\right)$ be the CDF of the $N\left(\mu, \sigma_d^2\right)$ evaluated at the point z_s, then we write the distance function as:

$$d(z_s, z_{s-1}) \propto \Phi(z_s; \mu, \sigma_d^2) - \Phi(z_{s-1}; \mu, \sigma_d^2). \tag{16}$$

Our MCMC algorithm involves sequentially drawing from $p(\theta|Data, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2, \gamma)$, $p(\alpha|Data, \theta, Q, \sigma_\eta^2, \mu, \sigma_d^2, \gamma)$, $p(Q^{-1}|Data, \theta, \alpha, \sigma_\eta^2, \mu, \sigma_d^2, \gamma)$, $p(\sigma_\eta^{-2}|Data, \theta, \alpha, \sigma_\eta^2, \mu, \sigma_d^2, \gamma)$, $p(\mu|Data, \theta, \alpha, Q, \sigma_\eta^2, \sigma_d^2, \gamma)$, $p(\sigma_d^2|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \gamma)$ and $p(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2)$. The first four of these posterior conditional distributions are standard (see the discussion surrounding the model defined by Eqs. (8)–(11)). Although note that some minor modifications of the algorithm of Kim et al. (1998) are required to draw from α due to the distance function entering the distribution of η_s. But these modifications are trivial since, by dividing (15) by $\sqrt{d(z_s, z_{s-1})}$ we obtain a state equation for volatilities that is in the same format as Kim et al. (1998). The final three of these posterior conditionals we discuss here.

For $p(\mu|Data, \theta, \alpha, Q, \sigma_d^2, \gamma)$ and $p(\sigma_d^2|Data, \theta, \alpha, Q, \mu, \gamma)$ we use Random Walk Chain Metropolis–Hastings algorithms (see, e.g. Chib and Greenberg, 1995). To be precise, if $p(\mu)$ is the prior, Bayes theorem implies

$$\begin{aligned} &p\left(\mu|Data, \theta, \alpha, Q, \sigma_d^2, \sigma_\eta^2, \gamma\right) \\ &\quad \propto p\left(Data, \theta, \alpha|\mu, \sigma_\eta^2, Q, \sigma_d^2, \gamma\right) p(\mu) \\ &\quad \propto p\left(Data|\theta, \mu, \alpha, \sigma_\eta^2, Q, \sigma_d^2, \gamma\right) \\ &\qquad p\left(\theta, \alpha|\mu, \alpha, \sigma_\eta^2, Q, \sigma_d^2, \gamma\right) p(\mu) \\ &\quad \propto p\left(\theta, \alpha|\mu, \sigma_\eta^2, Q, \sigma_d^2, \gamma\right) p(\mu), \end{aligned} \tag{17}$$

where the last line arises since, conditional on θ and α, the data provides no additional information about μ. Thus, to evaluate (17) at a point we need only evaluate $p(\theta, \alpha|\mu, \gamma)$, which by (13) and (15) is Normal, and the prior, $p(\mu)$. In our empirical work, we use a Normal prior and thus $\mu \sim N\left(\underline{\mu}, \underline{V}_\mu\right)$.

The Random Walk Chain Metropolis–Hastings algorithm generates candidate draws, μ^*, according to:

$$\mu^* = \mu^{(r-1)} + \zeta, \tag{18}$$

where $(r-1)$ superscripts (such as $\mu^{(r-1)}$) denote the $(r-1)$th draw from the algorithm and ζ is the increment random variable. We let $\zeta \sim N(0, c)$ and choose c to yield an average acceptance probability of roughly 0.50. Candidate draws are accepted with probability:

$$\begin{aligned} &A\left(\mu^{(r-1)}, \mu^*\right) \\ &= \min\left[\frac{p\left(\mu^*|\theta^{(r-1)}, \alpha^{(r-1)}, Q^{(r-1)}, \sigma_d^{2(r-1)}, \sigma_\eta^{2(r-1)}, \gamma^{(r-1)}\right)}{p\left(\mu^{(r-1)}|\theta^{(r-1)}, \alpha^{(r-1)}, Q^{(r-1)}, \sigma_d^{2(r-1)}, \sigma_\eta^{2(r-1)}, \gamma^{(r-1)}\right)}, 1\right], \end{aligned} \tag{19}$$

which can be evaluated using (17).

The next block in the MCMC algorithm involves a Random Walk Chain Metropolis–Hastings algorithm for σ_d^2. Since σ_d^2 is a positive random variable, we parameterize in terms of $\log\left(\sigma_d^2\right)$. Thus, the Random Walk Chain Metropolis–Hastings algorithm proceeds analogously to (18) and (19), the only difference is that we replace (18) by:

$$\log\left(\sigma_d^{2*}\right) = \log\left(\sigma_d^{2(r-1)}\right) + \zeta.$$

The acceptance probability is of a similar form as the algorithm for μ except that (17)–(19) are replaced with formula involving $\log\left(\sigma_d^2\right)$ instead of μ. To be precise, using similar steps as in (17) we have:

$$\begin{aligned} &p\left(\sigma_d^2|Data, \theta, \alpha, Q, \mu, \sigma_\eta^2, \gamma\right) \\ &\quad \propto p\left(\theta, \alpha|\mu, \sigma_\eta^2, Q, \sigma_d^2, \gamma\right) p\left(\sigma_d^2\right), \end{aligned} \tag{20}$$

where we take the prior, $p\left(\sigma_d^{-2}\right)$, to be $G\left(\underline{\mu}_d, \underline{\nu}_d\right)$ where this denotes the Gamma distribution with mean $\underline{\mu}_d$ and degrees of freedom $\underline{\nu}_d$. The change-of-variable term can be used to $p\left(\ln\left(\sigma_d^2\right)|Data, \theta, \alpha, Q, \mu, \sigma_\eta^2, \gamma\right)$ which involves the multiplication of (20) by the usual Jacobian term.

Finally, γ can be treated in several ways. γ can be interpreted as a model indicator and, thus, any of the standard approaches for calculating posterior model probabilities (or averaging across

arguments. For instance, $p\left(Q^{-1}|Data, \theta, \alpha, \sigma_\eta^2\right)$ does not depend on σ_η^2 and we could have written this conditional posterior as $p\left(Q^{-1}|Data, \theta, \alpha\right)$.

[14] The non-Bayesian equivalent of this need for prior information is the necessity of initializing the Kalman filter.

models) can be employed. If the number of configurations γ can take is fairly small, then the marginal likelihood can be calculated for each model using standard methods based on MCMC algorithms (e.g. Chib and Jeliazkov, 2001 or Gelfand and Dey, 1994). In terms of the notation of this paper, such an approach can be used to directly evaluate $p(\gamma|Data)$ (as opposed to drawing from $p\left(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2\right)$). Alternatively, if the computational cost of these methods of marginal likelihood calculation is high, then the researcher can use various approximations for the marginal likelihood (e.g. the Bayesian information criterion or the Laplace approximation of Tierney and Kadane (1986)). Yet another alternative is to use an algorithm which draws from model and parameter space jointly (i.e. in our context, this means drawing directly from $p\left(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2\right)$). Examples of such algorithms include Carlin and Chib (1995) and Green (1995). Carlin and Louis (2000) chapter 6 offers a useful overview of such algorithms. In this paper, we use simply using posterior expectation of the Bayesian information criterion to approximate the log of the marginal likelihood in each model. Note that, conditional on γ, μ, α and σ_d^2 (parameters which can be integrated out in the MCMC algorithm) our models are state space models and, thus, the prediction error decomposition form for the likelihood function (see, e.g., Harvey, 1989, page 91) allows for its easy evaluation using output from the Kalman filter.

In our empirical work with a TVP extension of an AR(p) model (one of our empirical exercises also includes a covariate p_t and its lags), we restrict the possible configurations of γ:

- $\gamma = (1, 2, 3, \ldots, T)'$, (i.e. $z_t = t$)
- data hypothetically ordered according to the jth lag of the dependent variable or (where relevant) other covariates (i.e. $z_t = y_{t-j}$ and $z_t = p_{t-j}$) for $j = 1, \ldots, p$.
- data hypothetically ordered to average of the lagged dependent variable or (where relevant) covariates over the last p periods (i.e. $z_t = \frac{\sum_{j=1}^{p} y_{t-j}}{p}$ and $z_t = \frac{\sum_{j=1}^{p} p_{t-j}}{p}$).

Thus, the number of possible choices for the index variable is $2p$ for our first empirical exercise and $4p - 1$ in our second empirical exercise (involving the additional covariate). Of course, in another empirical exercise other sets of configurations would be possible (and if the number of observations is small, the researcher may even wish to leave γ unrestricted).

4. Empirical work

4.1. The prior

In our work with artificial and real data a prior is required. Most of the parameters in our set of models are familiar ones in the time varying parameter and stochastic volatility literature (e.g. initial AR coefficients and error variance) and we will just use standard weakly informative priors for these. Of course, in more substantive empirical exercises, the researcher may wish to use noninformative priors, carry out a prior sensitivity analysis and/or adopt an objective prior approach (e.g. the training sample approach of O'Hagan (1995)). However, our distance function is new and so it is worthwhile to talk about its prior. This also allows us to understand the properties of the distance function.

Our distance function is given in (16) which depends on the parameters μ and σ_d^{-2}. Any prior is possible for these parameters and, particularly as we are using a Metropolis–Hastings algorithm, the researcher is unrestricted in prior choice. Here we assume the prior for μ to be $N\left(\underline{\mu}, \underline{V}_\mu\right)$ and the prior for σ_d^{-2} to be $G\left(\underline{\mu}_d, \underline{\nu}_d\right)$. Intuitively, by shifting μ around we can accommodate larger coefficient shifts in the region near μ. Small values of σ_d^2 are consistent with rapid changes in coefficients in the region near μ, whereas larger values of σ_d^2 are consistent with gradual evolution of coefficients (i.e. approaching a standard TVP model). In most cases, it will be desirable to have a prior which allows for all these possibilities. In all our empirical work below, we set $\underline{\mu} = 0$, $\underline{V}_\mu = 0.5$, $\underline{\mu}_d = 5$ and $\underline{\nu}_d = 5$. In the remainder of this subsection, we will discuss the implications of these choices. Remember that $\underline{\mu}_d$ is the mean of the precision, σ_d^{-2}, not the variance and, hence, this prior is allocating a great deal of weight to values of σ_d^2 less than one.

The distance function depends not only on μ and σ_d^{-2}, but also on z_s and z_{s-1}. In this subsection, we take $z_t = t$ for $t = 1, \ldots, T$. To aid in interpretation, we always standardize our variables so that they have mean zero and standard deviation one. This means, in particular, that our index variable will typically have almost all of its observations in the interval $[-2, 2]$. Furthermore, our distance function is standardized so as to have mean 1.[15] Hence, we have:

$$d(z_s, z_{s-1}) = \frac{\Phi(z_s; \mu, \sigma_d^2) - \Phi(z_{s-1}; \mu, \sigma_d^2).}{\frac{1}{T}\sum_{i=1}^{T}\left[\Phi(z_i; \mu, \sigma_d^2) - \Phi(z_{i-1}; \mu, \sigma_d^2)\right].} \tag{21}$$

The upper left-hand panel of Fig. 1 presents the prior mean of the distance function for every value of z_s. This can be seen to have an inverted-U shape but be relatively flat (i.e. distance between observations in the middle of the sample is less than four times as big as distance function at the very beginning or end of sample). However, this prior mean averages over many values for the parameters. The prior standard deviations accompanying these prior mean are roughly one, suggesting that wide deviations from the prior mean are possible.

The remaining three panels of Fig. 1 plot three particular distance functions arising from specific values of μ and σ_d^2. In particular, we have constructed the distance function for: (i) $\mu = -0.5$ and $\sigma_d^2 = 0.01$, (ii) $\mu = 1.6$ and $\sigma_d^2 = 0.5$ and (iii) $\mu = 0$ and $\sigma_d^2 = 2$. Note that all of these parameter configurations are in areas of appreciable prior probability using our prior. The upper right-hand panel of Fig. 1 shows how our parametric form for the distance function can accommodate an abrupt break in the first half of the sample. Note that, away from the mode, the distance function falls away to virtually zero (indicating constancy of coefficients) very quickly. The lower right-hand panel of Fig. 1 indicates that quite flat distance functions can be accommodated (consistent with a TVP model). The lower left-hand panel serves to illustrate that yet other types of behavior are consistent with our prior. In this case, we have a mode at the end of the sample, indicating increasing volatility in coefficients over time.

In sum, our functional form for the distance function and the prior we have chosen are extremely flexible, able to accompany a wide range of properties, including abrupt or gradual breaks in coefficients (at any point in the sample) or more gradual evolution of coefficients consistent with a TVP model.

4.2. Artificial data

Before proceeding to empirical work with real data, we begin by illustrating some of the aspects of our approach using artificial

[15] We find that some form of standardization produces more efficiency in our MCMC algorithm. Eq. (21) presents one simple standardization with the advantage it pins down the average value of the distance function. In some cases researchers might want to use a standardization that is not dependent on the particular sample. This can be done simply by choosing a normalization term based a minimum value of the index variable of -3, a maximum value of $+3$ and, for example, a divisor of $2T$.

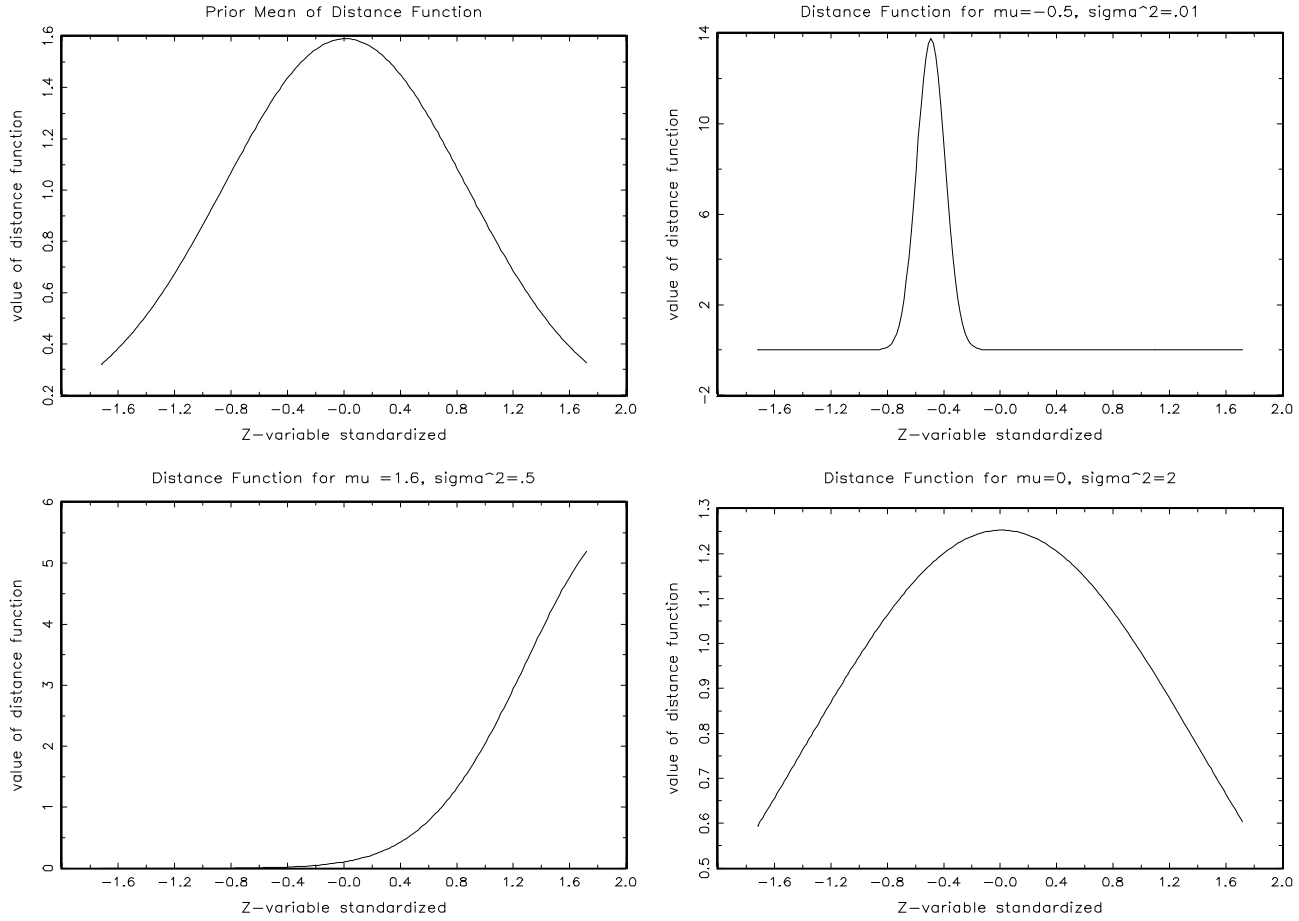

Fig. 1. Prior properties of distance function.

data. All the artificial data sets are generated from:

$$y_t = [1 - F(z_t)]\,\rho_1 y_{t-1} + F(z_t)\,\rho_2 y_{t-1} + \sigma_\varepsilon \varepsilon_t, \tag{22}$$

for $t = 1, \ldots, 200$ where ε_t is i.i.d. $N(0, 1)$. The coefficients evolve according to:

$$F(z_t) = \frac{1}{1 + \exp(-az_t)}. \tag{23}$$

For $z_t = y_{t-1}$ we have the familiar STAR model. In general, depending on a, we have a data generating process (DGP) where the AR(1) coefficient shifts either gradually or abruptly from one value to another. We set $a = 10$, $\rho_1 = 0$, $\rho_2 = 0.5$ and $\sigma_\varepsilon = 0.01$. However (as with all our variables in our empirical work), we then standardize the data to have mean zero and variance one.

We begin by focussing on the benefits brought by the addition of the distance function and, thus, do not consider hypothetical reorderings the data at this stage. Thus our first data set involves the choices above plus $z_t = t$ for $t = 1, \ldots, 200$. Furthermore, we focus on evolution of the AR coefficient and, thus, assume homoskedasticity (i.e. we do not allow for stochastic volatility as given in Eqs. (14) and (15)). Thus, our model depends on the parameters $\mu, \sigma_d^2, \sigma_\varepsilon^2, Q$ and θ where θ is the vector of AR coefficients. As described above, we separate out θ_0 and treat this as a regression effect. We use a weakly informative prior, $\theta_0 \sim N\left(0, 0.7^2\right)$ which reflects a (weak) prior belief in the stationarity. We assume σ_ε^{-2} is $G\left(\frac{1}{s^2}, 2\right)$ where s^2 is the OLS estimate of σ^2 in an AR(1) model. Thus we have a proper, but relatively diffuse prior centered around the comparable OLS quantity. Following standard practice we elicit our prior in terms of error precision matrices and assume $Q^{-1} \sim W\left(2, \frac{10}{2}\right)$ where $W(\nu, H)$ denotes the Wishart distribution with mean elements νH and degrees of freedom ν. This is a quite dispersed prior. In terms of Q (the variance in the state equation) it is centered approximately over 0.1 and, thus, we are allowing for everything from very small to moderately large shifts in the AR coefficients in each period (remember that the distance function has mean one as described in Eq. (21)). The priors for μ and σ_d^2 were described in the previous subsection.

Using the methods of posterior simulation described above, with $X_t = y_{t-1}$ and index definition variable simply being the ordering (i.e. $1, 2, \ldots, T$), we can obtain posterior properties of any of the model parameters (or functions thereof). Figs. 2 and 3 graph some aspects of particular interest. Fig. 2 plots the true value of the AR coefficient used to generate the data. It also plots the OLS estimate of the AR(1) model (without intercept) using this data. Finally, it plots the posterior mean of θ estimated using our model. It can be seen that, even with a relatively small number of observations and small break in the AR coefficient, our model tracks the true value fairly well (except at the very beginning of the sample). The fairly abrupt break in the AR coefficient in the DGP is matched reasonably well by our model. Presumably the distance function gets much larger near where the break is in order to accommodate it. Fig. 3 confirms the contention of the preceding sentence. It plots the posterior mean of the distance function. It can be seen that the value of the distance function becomes larger near where the break is (to allow larger changes in the AR coefficient in this region).

To illustrate the benefits of considering hypothetical reorderings, Data Set 2 is generated from (22) and (23) from a STAR specification where the data switches between a "recessionary" regime with transitory dynamics to an "expansionary" regime with more

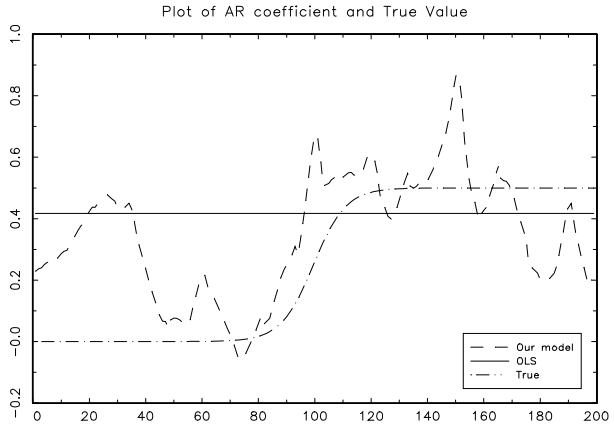

Fig. 2. The AR coefficient for data set 1.

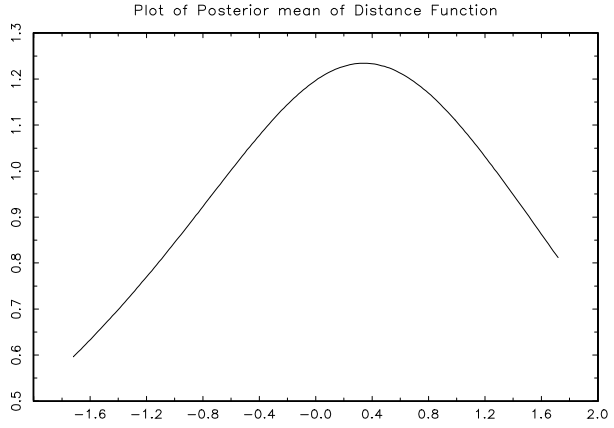

Fig. 3. The distance function for data set 1.

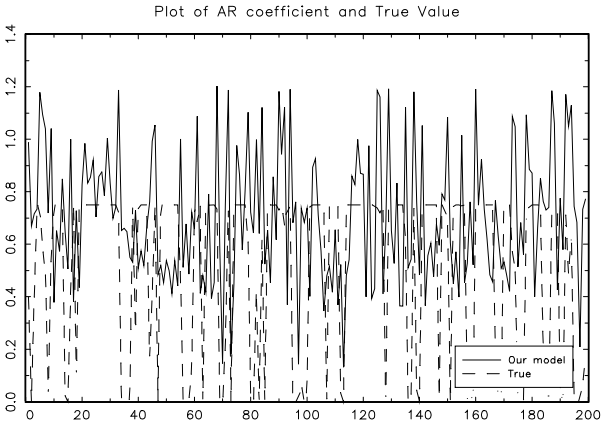

Fig. 4. The AR coefficient for data set 2 (our model).

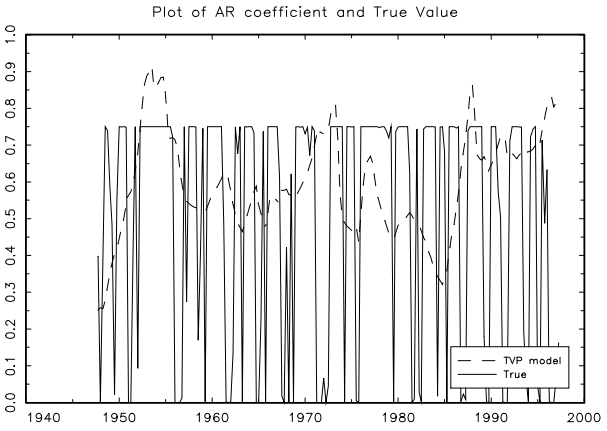

Fig. 5. The AR coefficient for data set 2 (TVP model).

persistent dynamics. In particular, we set $z_t = y_{t-1}$, $a = 1000$, $\rho_1 = 0$, $\rho_2 = 0.75$ and $\sigma_\varepsilon = 0.01$. We then standardize the data to have mean zero and variance one. We consider the set of orderings of the data described at the end of Section 3. Thus, since $p = 1$, we order the data in normal time ordering and according to y_{t-1}. All other aspects of the data generating process and prior are the same as used for Data Set 1.

The model we propose is particularly useful in nonparametric contexts: where the researcher is unsure of the form of the nonlinear time series model used to generate the data. Our model, with Data Set 2, is designed to reflect the (common) case where the researcher thinks there might be structural breaks or TAR-type regime-switching (or both), but is not sure which. With this data set, our model does well in picking out the correct form of regime switching. In particular, we find that there is a 94.3% probability that the (correct) model with data ordered by y_{t-1} is the correct one and only a 5.7% probability that the normal time ordering is appropriate. These weights are used when we average over the two orderings when calculating the following results.

Fig. 4 plots the true value of the AR coefficient used to generate the data along with its posterior mean calculated using our model (averaged over both definitions for the index variable). And, although it does have some trouble picking up all the many regime-switches in the DGP and sometimes wanders into the nonstationary region (something which could be avoided through use of a more informative prior), it does do moderately well at matching many of the regime-switches. As a benchmark for comparison, we also estimate a standard TVP model. This is a special case of our model with the distance function defined to always be one. We use the same prior for the TVP parameters as for the parameters of our model. Fig. 5 is comparable to Fig. 4. Clearly it can be seen that the TVP model is not picking up the regime-switches nearly as well as our model. We note also that the BICs for our model (with data ordered according to y_{t-1}), our model (with data in natural time ordering) and the TVP model are: -48.627, -51.435 and -51.404. Of course, a correct parametric model would beat our model. But in the common case where it is not clear which parametric model to use, and the researcher wishes to use a flexible approach (e.g. the TVP model), our approach does seem to be promising. As yet another metric of the performance of our approach, we calculated the correlation between the actual data and: (i) the fitted OLS line, (ii) the fitted values provided by the TVP model (evaluated at posterior means) and (iii) the fitted values provided by our model (evaluated at posterior means), we obtain values of 0.666, 0.718 and 0.739, suggestive of moderate improvements in fit of using our approach.

Finally, we present the posterior mean of the distance function. Note that when we are considering the reordering by y_{t-1}, we estimate the distance function using this ordering. Since our approach involves multiple choices for z_t, Fig. 6 transforms back to the natural time ordering. This accounts for the irregular shape of Fig. 6. Note that large values of the distance function tend to be associated with times where regime-switches occur, whereas smaller values tend to be associated with times where the AR coefficient is unchanging.

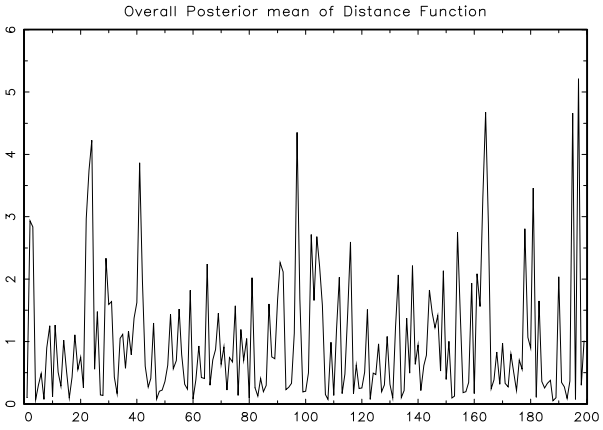

Fig. 6. The distance function for data set 2.

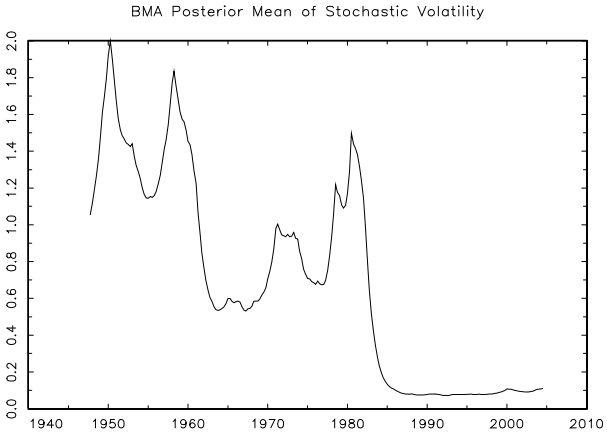

Fig. 7. Stochastic volatility (univariate example).

4.3. Empirical illustrations using real GDP growth

4.3.1. Univariate properties of GDP growth

There are many applications which investigate nonlinearities or structural breaks in real GDP growth. For instance, Beaudry and Koop (1993) and Potter (1995) are early papers which investigate nonlinearities in the conditional mean, $E(y_t|I_{t-1})$ where I_{t-1} denotes data information through time $t-1$, using models where AR dynamics change over the business cycle. More recently, there has been interest in the volatility of US real activity and the question of whether it has decreased over time. This finding is sometimes referred to as the Great Moderation of the business cycle. For instance, Kim et al. (2004) investigate breaks in the volatility of various measures of aggregate activity. For most of the measures they consider, they find strong evidence of an abrupt break in the early 1980s. Stock and Watson (2002) find similar evidence for a change in volatility, but find the decline to have been more gradual, a thesis also put forward by Blanchard and Simon (2001). Thus, using measures of real output, a wide variety of regime-switching and structural break models for the conditional mean and conditional variance have been used. Some allow for gradual change between regimes, others are more abrupt. Our model will nest all these possibilities.

We consider extensions of the AR(p) model and let X_t contain an intercept plus p lags of y_t for $p = 1, 2, \ldots, 5$. For Q, μ, σ_d^2 we use the same weakly informative prior as in the artificial data section. To incorporate a noninformative prior for the intercept, we extend the weakly informative prior used previously to

$$\theta_0 \sim N\left(0_{p+1}, \begin{bmatrix} 10^{10} & 0 & 0 & 0 \\ 0 & 0.7^2 & 0 & 0 \\ 0 & 0 & \ldots & \ldots \\ 0 & 0 & \ldots & 0.7^2 \end{bmatrix}\right)$$

which still reflects a (weak) prior belief that the AR coefficients at each point in time lie in the non-explosive region of the parameter space. For the stochastic volatility component we assume σ_η^{-2} to be $G(10, 50)$.

Consider first the issue of lag length. For any choice of p, we can sum over all index variables to calculate $\Pr(p = j|Data)$ for $j = 1, \ldots, 5$. These probabilities are 0.003, 0.024, 0.640, 0.331 and 0.002, respectively. Thus, there is uncertainty over lag length, but $p = 3$ receives the most support. The results in the remainder of this section do Bayesian model averaging over these five choices for lag length.

Regardless of whether we average over lag length or simply choose a particular value for p, with our model we find strong evidence in favor of a decrease in the volatility of GDP growth. Indeed this evidence is so strong, that it swamps any evidence for the regime-switching behavior in the AR coefficients found, e.g., by Beaudry and Koop (1993) and Potter (1995). Note first that the results discussed below are averaged over all of our different choices for z_t but, in practice, the probability that $z_t = 1$ is over 99% and, accordingly, results are very similar to those found if we had selected a single model and worked with observations in the natural time ordering. Fig. 7 plots the posterior mean of the error variance and it can be seen that there is a big drop in volatility around 1983. Our findings relating to volatility of GDP growth are not surprising given previous results starting with McConnell and Perez (2000) (see also the similar figure in Koop and Potter, 2007). There is some evidence that volatility started to decline in the 1950s but this decline was reversed starting around 1970. A single break model (by construction) could not show this kind of pattern.

Remember that, in our model, the breaks in the error variances and AR coefficients are assumed to occur at the same time. This accounts for the fact that we are not finding strong evidence of breaks or regime-switching behavior in the AR coefficients. The break in volatility is so strong that it receives overwhelming support relative to other models. It is straightforward to extend our model to allow for different orderings and distance functions to hold for the AR coefficients and the error variances and we do so in the context of our next empirical illustration.

Our model clearly outperforms a benchmark AR(p) and, hence, we will not present results for an AR(p). A preferable benchmark, which a researcher might use when agnostically approaching a data set which is suspected to have structural change, would be the TVP model with stochastic volatility given in (8)–(11) or, equivalently, our model with $\gamma = (1, \ldots, T)'$ and $d(z_s, z_{s-1}) = 1$ for all observations. Given the fact that our model yields strong evidence in favor of $\gamma = (1, \ldots, T)'$, it is not surprising that we find results to be similar between our approach and the TVP model. Fig. 8 plots the posterior mean of the distance function using our model. It can be seen to be inverted-U shape, thus smoothing observations at the beginning and end of the sample more than observations at the middle. This accounts for the differences between our model and the TVP model that can be seen if one compares Figs. 7 and 9 (note the difference in scaling of the y-axis). In this data set (and for macroeconomic policy), the patterns in the error variance seem to be the most crucial. Note that, relative to a researcher using a TVP model, we are finding the same general pattern of volatility, but lower and less erratic volatility in the 1950s and a smoother pattern after the 1983 break in volatility. Furthermore, our model does fit a bit better. As a rough metric we calculated the correlation between the expected value of the dependent variable and the observed dependent variable. This is

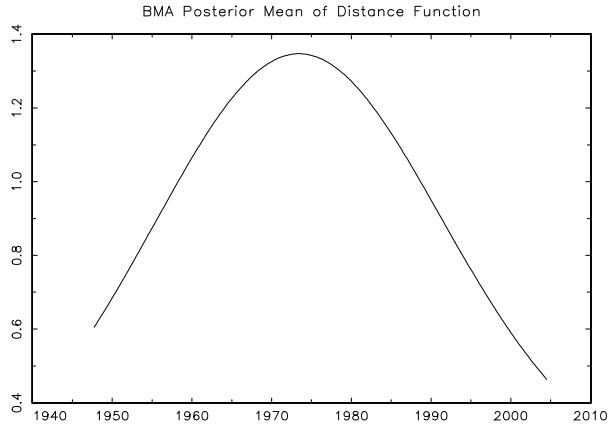

Fig. 8. Distance function (univariate example).

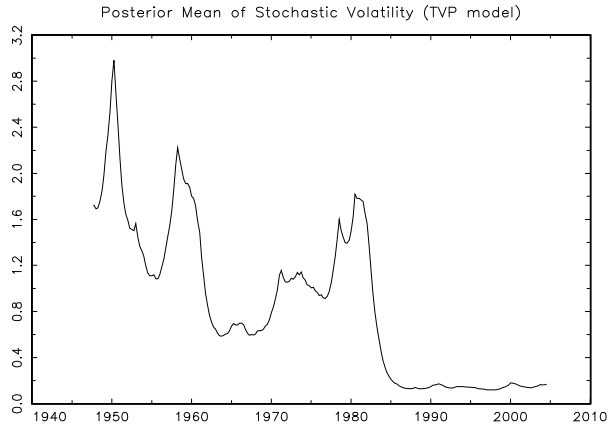

Fig. 9. Stochastic volatility for TVP model (univariate example).

0.787 for our model, but 0.681 for the TVP model (and only 0.342 for an AR(2)).

4.4. The oil price and GDP growth

There is a large literature on the effects of oil price changes on GDP growth. Hamilton (2003) is an important contribution to this literature and his list of references cites much previous work. Hamilton provides a compelling argument that functional form issues are important when seeking to understand the relationship between oil prices and GDP. If y_t is GDP growth and p_t the percentage change in the oil price, then Hamilton begins with a linear specification with four lags of both variables:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_4 y_{t-4} + \beta_5 p_{t-1} + \cdots + \beta_8 p_{t-4} + \sigma_\varepsilon \varepsilon_t,$$

before presenting convincing evidence of departures from linearity using a flexible approach to nonlinear inference developed in Hamilton (2001). Note that this approach focusses on nonlinearities in the conditional mean, $E(y_t|I_{t-1})$ (with the conditional variance, $var(y_t|I_{t-1})$, being of less interest). Below we will refer to the parameters which characterize this conditional mean as regression coefficients (these are labelled θ_t in Eq. (13)).

In this section, inspired by Hamilton (2003), we illustrate our methods using an updated data set where the real GDP and oil price series used to construct growth rates run from 1947Q1 through 2006Q4.[16] Our set of explanatory variables (i.e. X_t in Eq. (8)), is an intercept, four lags of GDP growth and four lags of oil price growth (i.e. log differences of the original GDP and oil price series).

In our previous empirical example, we showed how the Great Moderation of the business cycle (i.e. the reduction in the volatility of GDP growth) dominates any possible effects in the regression coefficients. The same thing happens here. That is, if we use the same setup as in the univariate GDP growth example (extended to allow for four lags of oil price growth), we find overwhelming evidence in favor of $\gamma = (1, \ldots, T)'$. In light of this (and in order to illustrate an empirically useful extension), in this section we present results where different hypothetical orderings and distance functions exist for the regression coefficients and the conditional variance, α_t (i.e. the state Eqs. (13) and (15) can be based on different orderings). That is, we now have a γ_1 (which controls ordering relating to θ_t) and γ_2 (for α_t). For γ_1, we use the same choices as before (defined at the end of Section 3) and add lagged oil price changes and long averages of them. For γ_2 we could use the same set of choices, but since the Great Moderation implies $\gamma_2 = (1, \ldots, T)'$ is so predominant, the results below just use this choice. For our modified state space model defined in (12)–(15), we now have two distance functions so that $v_s \sim N\left[0, \sqrt{d_1(z_s, z_{s-1})}Q\right]$ and $\eta_s \sim N\left[0, \sqrt{d_2(z_s, z_{s-1})}\sigma_\eta^2\right]$ (and the index variables in the two distance functions are potentially different from one another).

For the parameters in common with the model used in the previous section, we use the same prior as specified there. With regards to the new parameters, the coefficients on the lagged oil price inflation variables have the same prior as those on the lagged GDP growth variables. For each of the two distance functions we use the same prior as we used for the single distance function in the previous example.

Table 2 presents some empirical results for each of our index variables controlling the ordering relating to the regression coefficients (i.e. different choices for γ_1). In contrast to our univariate results, there is now a certain degree of uncertainty over orderings. Note that the models defined by different choices for γ_1 and γ_2 have the same number of parameters and, hence, BICs will simply be proportional to log-likelihoods. Accordingly, we simply present log-likelihoods (calculated as discussed in Section 3) in Table 2. As discussed previously, BICs can be used to approximate log marginal likelihoods and, thus, can be used to calculate posterior model probabilities associated with each choice of γ_1 and γ_2. Less than 2% of the probability is associated with the choice which implies the regression coefficients evolve according to a conventional TVP model. However, most of the probability is attached to models where the hypothetical reordering observations is with respect to past oil price inflation. The most probable ordering is based on the index variable $z_t = \frac{\sum_{j=1}^{4} p_{t-j}}{4}$. Thus, we are finding most support for a model which is similar to a STAR model where the different regimes are triggered by the average increase in oil prices over the past year. If we were to treat each choice for γ_2 as defining a model, then any conventional Bayesian or non-Bayesian hypothesis testing/model selection procedure would choose the model based on the index variable $z_t = \frac{\sum_{j=1}^{4} p_{t-j}}{4}$.

To provide a broader comparison with conventional models, note that the log-likelihoods (evaluated at the MLE) for the AR(2) and TVP (with stochastic volatility) models are −97.25 and −50.80, respectively. Clearly, our model is performing massively better than the AR(2). Furthermore, since the highest log-likelihood in Table 2 is −43.73, we are also obtaining substantial improvements relative to a standard TVP (with stochastic volatility) model.

We remind the reader that, with regards to the measurement equation error variance, we are assuming a stochastic volatility

[16] Following Hamilton (2003), we use the nominal crude oil producer price index.

Table 2
Probability of each ordering and posterior properties of key parameters (numbers in parentheses are posterior standard deviations).

Index for cond. mean	Prob.	Log-likelihood	Dist. function cond. mean		Dist. function cond. variance	
			$E(\mu)$	$E(\sigma_d^2)$	$E(\mu)$	$E(\sigma_d^2)$
t	0.015	−47.69	−0.417 (0.503)	5.874 (7.222)	−0.299 (0.624)	4.637 (4.795)
y_{t-1}	0.000	−54.67	−0.278 (0.510)	10.721 (24.048)	−0.471 (0.601)	3.858 (4.447)
y_{t-2}	0.000	−54.25	−0.160 (0.598)	3.162 (2.478)	−0.366 (0.427)	3.984 (3.411)
y_{t-3}	0.000	−53.27	0.102 (0.433)	1.804 (0.880)	−0.313 (0.542)	5.987 (5.319)
y_{t-4}	0.000	−51.14	0.161 (0.555)	1.902 (0.901)	−0.448 (0.484)	4.287 (4.710)
$\frac{\sum_{j=1}^{2} y_{t-j}}{2}$	0.000	−54.46	0.031 (0.573)	3.557 (2.408)	−0.319 (0.446)	3.415 (3.985)
$\frac{\sum_{j=1}^{3} y_{t-j}}{3}$	0.003	−49.46	0.180 (0.424)	1.227 (0.487)	0.277 (0.586)	5.458 (4.266)
$\frac{\sum_{j=1}^{4} y_{t-j}}{4}$	0.010	−48.13	0.074 (0.493)	5.193 (5.304)	−0.424 (0.828)	3.487 (1.798)
p_{t-1}	0.000	−54.20	0.624 (0.383)	0.866 (0.177)	−0.111 (0.765)	2.610 (2.677)
p_{t-2}	0.000	−53.40	0.644 (0.345)	1.392 (0.692)	−0.759 (0.658)	6.058 (8.863)
p_{t-3}	0.003	−49.43	0.656 (0.663)	13.479 (13.971)	−0.217 (0.410)	2.761 (2.328)
p_{t-4}	0.000	−55.59	0.676 (0.359)	1.323 (0.490)	−0.247 (0.477)	3.678 (5.680)
$\frac{\sum_{j=1}^{2} p_{t-j}}{2}$	0.000	−54.40	0.493 (0.427)	2.520 (0.879)	−0.304 (0.518)	7.410 (9.347)
$\frac{\sum_{j=1}^{3} p_{t-j}}{3}$	0.173	−45.26	0.857 (0.409)	2.406 (1.312)	−0.081 (0.571)	4.470 (4.720)
$\frac{\sum_{j=1}^{4} p_{t-j}}{4}$	0.797	−43.73	0.914 (0.561)	2.177 (0.857)	−0.282 (0.638)	4.268 (4.667)

specification and simply setting $\gamma_2 = (1, \ldots, T)'$ (although the addition of our distance function means we are not exactly in a conventional stochastic volatility framework). As in our univariate GDP growth example, we find strong evidence in favor of this choice. However, since results in this regard are very similar to those in the univariate example (e.g. a plot of the posterior mean of stochastic volatility looks very similar to Fig. 7), we will not discuss them further.

Table 2 also reports the posterior means and standard deviations of the parameters characterizing the distance functions. The most important pattern in these is the fact that the posterior means for the variances in the two distance functions (i.e. the σ_d^2s) both are quite large. Remember that small values of this parameter will imply abrupt regime-switches (e.g. characteristic of TAR or standard structural break models) whereas large values of this parameter imply gradual evolution between regimes (characteristic of STAR or TVP models). We are finding support for the latter case.

Thus far we have said little of the economic implications of our findings. Fig. 10 illustrates how this can be done. In the original linear model, a rough measure of the effect of changes in the oil price on GDP growth is the sum of the coefficients on the lags of the oil inflation variables (i.e. $\beta_5 + \cdots + \beta_8$). In our preferred specification (i.e. the one where the index variable is the average change in oil prices over the last year), this measure will depend on the index variable. Fig. 10 plots the posterior mean of this measure against the index variable. Remember that the index variable has been normalized so that a value of 0 implies the mean value of oil price changes over the past year, 1 implies oil price changes over the past year one standard deviation above the mean, etc.[17] A general pattern in this figure is that drops in the oil price have a smaller marginal effect on output than rises (as has been found by Hamilton). But the pattern is quite erratic (and a large exception to it occurs around 1.0). Although it is true that very large positive oil price shocks (e.g. where the index is about 2) have the largest negative effects on GDP growth, the relationship between the index variable and this measure of the effect of an oil shock is highly non-monotonic.

Hamilton (2003), using a shorter data set, has shown that certain nonlinear transformations of oil prices have predictive power for US GDP growth. Models estimated on the earlier data predict larger effects from the recent price increases than what has recently been observed. In terms of the index variable used to produce Figs. 10 and 6 of the extra 21 observations we have added relative to Hamilton are between 0.8 and 1.2 (i.e. in the region where the marginal effect of oil prices on GDP growth is near zero). Since these changes had little effect on GDP growth, this probably explains why models estimated on the shorter data sets imply somewhat larger effects of moderate oil price rises on GDP growth.

To illustrate this point more clearly, Fig. 11 is the same as Fig. 10, but is produced using data through the end of 1997. We choose this date since it roughly corresponds with the trough in oil prices. Since then (with some exceptions) the oil price has been rising.

[17] Note that there are very few observations of the index at values greater than 2 in absolute value and the distance between them is sometimes large. This accounts for the flat regions near the boundaries of Fig. 10.

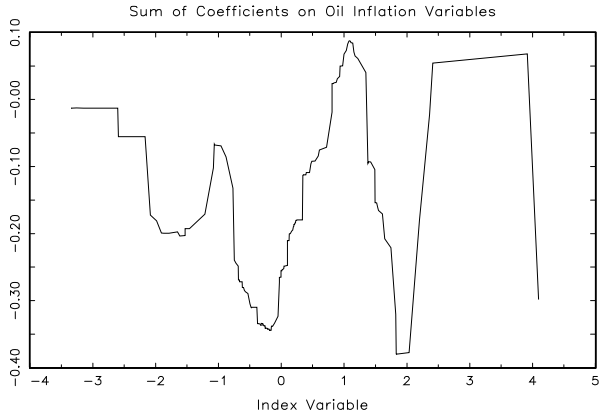

Fig. 10. A measure of the effect of an oil shock for the preferred ordering.

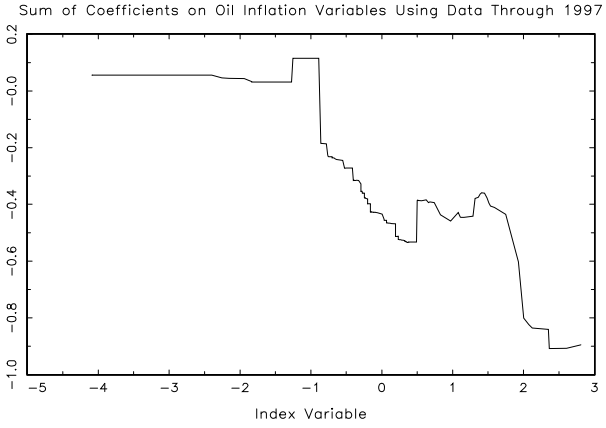

Fig. 11. A measure of the effect of an oil shock for the preferred ordering using data through 1997.

Note that Fig. 11 is more consistent with Hamilton's story that positive oil shocks have larger marginal effects on output than do negative oil shocks.

The empirical illustration shows the advantages and flexibility of our approach. Clearly, there are strong nonlinearities in the oil price–GDP growth relationship and, hence, our approach is finding things that a linear model could not. However, conventional nonlinear and nonparametric approaches typically involve making trying a single (or, at most, a few) choices analogous to our choice of an index variable. But there are so many possible choices of index, that conventional approaches risk misspecification problems (i.e. if a wide range of possibilities are not tried) or sequential testing problems (i.e. if a wide range of possibilities are tried and numerous tests are done to try and select a single index out of the myriad of possibilities).

5. Conclusion

For researchers working with macroeconomic and financial data, there is great interest in investigating whether structural breaks and/or regime-switching behavior occurs (in the conditional mean, $E(y_t|I_{t-1})$, and/or the conditional variance, $var(y_t|I_{t-1})$). In this paper, we have developed an extremely flexible parametric model which can accommodate each of these choices. We feel our model is an attractive one due to its simplicity. That is, it adds two simple concepts to a standard state space framework. These ideas are hypothetical reordering and distance. By imagining a range of orderings of the evolution of the parameter vectors, we can accommodate a wide variety of nonlinear time series models, including regime-switching and structural breaks. By allowing the state equation variances to depend on the distance between observations, we can accommodate a much wider variety of ways that our parameters can evolve, including everything from abrupt change models (e.g. threshold autoregressive models or structural break models such as that of Bai and Perron (1998)) to those which allow gradual evolution of parameters (e.g. smooth transition autoregressive models or TVP models such as that of Primiceri (2005)). In short, our model will nest virtually every popular model in the regime-switching and structural break literatures.

Moreover, because we retain the state space framework, Bayesian econometric methods and, especially, posterior simulation, are relatively straightforward, drawing on the existing literature. Our work with artificial and real data show the advantages of our approach.

References

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.
Beaudry, P., Koop, G., 1993. Do recessions permanently change output? Journal of Monetary Economics 31, 149–164.
Bec, F., Rahbek, A., Shephard, N., 2008. The ACR model: a multivariate dynamic mixture autoregression. Oxford Bulletin of Economics and Statistics 70, 583–618.
Blanchard, O., Simon, J., 2001. The long and large decline in US output volatility. Brookings Papers on Economic Activity 1, 135–174.
Boivin, J., Giannoni, M., 2006. Has monetary policy become more effective? Review of Economics and Statistics 88, 445–462.
Carlin, B., Chib, S., 1995. Bayesian model choice via Markov Chain Monte Carlo methods. Journal of the Royal Statistical Societ, Series B 57, 473–484.
Carlin, B., Louis, T., 2000. Bayes and Empirical Bayes Methods for Data Analysis, second ed. Chapman and Hall, Boca Raton.
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., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49, 327–335.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96, 270–281.
Chipman, H., George, E., McCulloch, R., 2001. The practical implementation of Bayesian model selection, manuscript. Available at http://gsbwww.uchicago.edu/fac/robert.mcculloch/research/papers/index.html.
Cogley, T., Sargent, T., 2001. Evolving post world war II inflation dynamics. NBER Macroeconomics Annual 16, 331–373.
Cogley, T., Sargent, T., 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. Review of Economic Dynamics 8, 262–302.
Cogley, T., Sargent, T., 2007. Inflation gap persistence in the US working paper. Available at http://www.econ.ucdavis.edu/faculty/twcogley/.
Durbin, J., Koopman, S., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.
Durbin, J., Koopman, S., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89, 603–616.
Fernandez, C., Osiewalski, J., Steel, M.F.J., 1997. On the use of panel data in stochastic frontier models with improper priors. Journal of Econometrics 79, 169–193.
Gelfand, A., Dey, D., 1994. Bayesian model choice: Asymptotics and exact calculations. Journal of the Royal Statistical Society Series B 56, 501–514.
Gerlach, R., Carter, C., Kohn, E., 2000. Efficient Bayesian inference in dynamic mixture models. Journal of the American Statistical Association 95, 819–828.
Giordani, P., Kohn, R., 2008. Efficient Bayesian inference for multiple change-point and mixture innovation models. Journal of Business and Economic Statistics 26, 66–77.
Giordani, P., Kohn, R., van Dijk, D., 2007. A unified approach to nonlinearity, outliers and structural breaks. Journal of Econometrics 137, 112–133.
Green, P., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.
Hamilton, J., 2001. A parametric approach to flexible nonlinear inference. Econometrica 69 (3), 537–562.
Hamilton, J., 2003. What is an oil shock? Journal of Econometrics 113, 363–398.
Harvey, A., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
Harvey, A., Koopman, S., 2000. Signal extraction and the formulation of unobserved components models. Econometrics Journal 3, 84–107.
Hoeting, J., Madigan, D., Raftery, A., Volinsky, C., 1999. Bayesian model averaging: a tutorial. Statistical Science 14, 382–417.
Kim, C., Nelson, C., Piger, J., 2004. The less volatile US economy: a Bayesian investigation of timing, breadth, and potential explanations. Journal of Business and Economic Statistics 22, 80–93.
Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.
Koop, G., 2003. Bayesian Econometrics. John Wiley, Chichester.

Koop, G., Poirier, D., 2004. Bayesian variants of some classical semiparametric regression techniques. Journal of Econometrics 123, 259–282.
Koop, G., Potter, S., 1999. Dynamic asymmetries in US unemployment. Journal of Business and Economic Statistics 17, 298–312.
Koop, G., Potter, S., 2001. Are apparent findings of nonlinearity due to structural instability in economic time series? The Econometrics Journal 4, 37–55.
Koop, G., Potter, S., 2006. The vector floor and ceiling model. In: Milas, C., Rothman, P., van Dijk, D. (Eds.), Nonlinear Time Series Analysis of the Business Cycle. In: Contributions to Economic Analysis, Elsevier (Chapter 4).
Koop, G., Potter, S., 2007. Estimation and forecasting in models with multiple breaks. Review of Economic Studies 74, 763–789.
Levin, A., Piger, J., 2008. Bayesian model selection for structural break models. Working paper. Available at SSRN: http://ssrn.com/abstract=1132463.
Lundbergh, S., Terasvirta, T., van Dijk, D., 2003. Time-varying smooth transition autoregressive models. Journal of Business and Economic Statistics 21, 104–121.
McConnell, M., Perez, G., 2000. Output fluctuations in the United States: what has changed since the early 1980s? American Economic Review 90, 1464–1476.
McCulloch, R., Tsay, E., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. Journal of the American Statistical Association 88, 968–978.
Potter, S., 1995. A nonlinear approach to US GNP. Journal of Applied Econometrics 10, 109–125.
O'Hagan, A., 1995. Fractional Bayes factors for model comparison. Journal of the Royal Statistical Society, Series B 57, 99–138.
Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72, 821–852.
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., Zha, T., 2006. Were there regime-switches in macroeconomic policy? American Economic Review 96, 54–81.
Stock, J., Watson, M., 2002. Has the business cycle changed and why? In: NBER Macroeconomic Annual.
Stock, J., Watson, M., 2007. Why has US inflation become harder to forecast? Journal of Money, Credit and Banking 39, 3–33.
Tierney, L., Kadane, J., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81, 82–86.
Tsay, R., 1989. Testing and modeling threshold autoregressive processes. Journal of the American Statistical Association 84, 231–240.
Wecker, W., Ansley, C., 1983. The signal extraction approach to nonlinear regression and spline smoothing. Journal of the American Statistical Association 78, 81–89.
Yatchew, A., 1998. Nonparametric regression techniques in economics. Journal of Economic Literature 36, 669–721.