Federal Reserve Bank of New York
Staff Reports

A Flexible Approach to Parametric Inference
in Nonlinear Time Series Models

Gary Koop
Simon Potter

A Flexible Approach to Parametric Inference in Nonlinear Time Series Models
Gary Koop and Simon Potter
Federal Reserve Bank of New York Staff Reports, no. 285
May 2007
JEL classification: C11, C22, E17

Abstract

Many structural break and regime-switching models have been used with macroeconomic and financial data. In this paper, we develop an extremely flexible parametric model that accommodates virtually any of these specifications—and does so in a simple way that allows for straightforward Bayesian inference. The basic idea underlying our model is that it adds two concepts to a standard state space framework. These ideas are ordering and distance. By ordering the data in different ways, we can accommodate a wide range of nonlinear time series models. By allowing the state equation variances to depend on the distance between observations, the parameters can evolve in a wide variety of ways, allowing for models that exhibit abrupt change as well as those that permit a gradual evolution of parameters. We show how our model will (approximately) nest almost every popular model in the regime-switching and structural break literatures. Bayesian econometric methods for inference in this model are developed. Because we stay within a state space framework, these methods are relatively straightforward and draw on the existing literature. We use artificial data to show the advantages of our approach and then provide two empirical illustrations involving the modeling of real GDP growth.

Key words: Bayesian, structural break, threshold autoregression, regime switching, state space model

Potter: Federal Reserve Bank of New York (e-mail: simon.potter@ny.frb.org). Koop: University of Strathclyde (e-mail: gary.koop@strath.ac.uk). The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

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 and Koop and Potter, 2006) develop regime-switching models to analyze similar issues relating to the effects of 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 lead 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, Terasvirta and van Dijk, 2003 and Rahbek and Shephard, 2002 and van Dijk, Giordani and Kohn, 2007). Our paper is in the spirit of this latter literature but, we argue, more flexible than previous work. We remain in a standard state space framework and, thus, textbook results for estimation, model comparison and prediction are immediately available.

The intuition underlying our modeling framework is based on two ideas: data reordering and distance between 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 if $y_{t-1} < \tau$ are different from those if $y_{t-1} \geq \tau$). 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 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 re-ordered data y_t^*. In general, a wide range of nonlinear time series models with regime-switches can be written as structural break models using reordered data.

The second idea our modeling framework draws on is distance between observations. This idea is incorporated in many nonparametric regression methods (see, e.g., Yatchew, 1998). For instance, 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 various nonparametric regression methods imply that they have similar values for $f(x)$. In a macroeconomic context such ideas are very useful if dynamic properties vary over the business cycle. For instance, in an autoregressive model we might want AR dynamics in all recessionary periods to be similar to one another. Thus, if y_{t-1} is last period's GDP growth, then two observations with similar negative values of y_{t-1} should have similar AR dynamics. A nonparametric time series regression model has such a feature, and our modeling framework has it as well.

In next section, we provide more detail on why combining these two ideas results in a modeling framework that is flexible and nests virtually every popular parametric structural

---

[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).

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 simple. 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, every model we consider is a state space model. 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 the third section 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. The fourth section 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 the switch between regimes.

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:

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

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

$$\rho_t = \rho_{t-1} + \sigma_v v_t \tag{2}$$

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 and Harvey and Koopman, 2000). Even without the extensions considered in this paper, state space models such as (1) and (2) are a flexible and powerful tool for time series analysis.

What we do in this paper is maintain the framework given by (1) and (2) (and, thus, still use standard MCMC algorithms), but extend it as follows:

$$y_t = \rho_{g(z_t)} x_t + \sigma_\varepsilon \varepsilon_t, \tag{3}$$

where

$$\rho_s = \rho_{s-1} + \sigma_v d\left(z_s^*, z_{s-1}^*\right) v_s, \tag{4}$$

and s indexes a reordering of time according to the function g of an *index variable* z_t (i.e. z_t is an observed exogenous variable such as a lagged dependent variable). To define g take the observations of the index variable z_t and place them in ascending order (using a * to denote the reordered values) as $z_1^* \leq z_2^* \leq .. \leq z_T^*$. Then $g(z_t)$ is the rank of z_t in the ordering. So, for instance, if $z_t = y_{t-1}$, and z_t has the tenth smallest value of y_{t-1}, then $g(z_t) = 10$. Formally, let the $T \times 1$ parameter γ with $\gamma_t = g(z_t)$ for $t = 1, ..., T$ be used to define the ordering of the data (e.g. $\gamma = (1, ..., T)'$ denotes data in the usual time series ordering, $\gamma = (T, T-1, .., 1)'$ denotes the data in reverse time ordering, etc.). $d\left(z_s^*, z_{s-1}^*\right)$ is a (non-negative) *distance function* measuring the distance between z_s^* and z_{s-1}^*. We treat γ and $d = \left(d\left(z_2^*, z_1^*\right), d\left(z_3^*, z_2^*\right), ..., d\left(z_T^*, z_{T-1}^*\right)\right)'$ as vectors of parameters (with hierarchical priors) and, thus, our methods allow for their data-based estimation.

Different reorderings of the data and different distance functions can be used to nest a wide range of common specifications as the following discussion make clear. To draw out the connections with existing nonlinear time series models most clearly, the following discussion assumes x_t contains a lag of the dependent variable.

2.1 The Role of the Distance Function if the Data is Not Reordered

Consider first, the sorts of models which result if the data are in standard time series order. In terms of our definitions, we have $\gamma = (1, ..., T)'$, $z_t = z_t^* = t$ for $t = 1, ..., T$ (i.e. the index variable is already in ascending order and, thus, no reordering of the data is required to define s or z_s^*). Thus, the model is:

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

where

$$\rho_t = \rho_{t-1} + \sigma_v d\left(t, t-1\right) v_s. \tag{6}$$

Two 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.

Now consider what happens if we let $d(t, t-1) = 1$ if $t = \tau$ and $d(t, t-1) = 0$ otherwise. In this case, $\rho_1 = ... \rho_{\tau-1}$ and $\rho_\tau = ... \rho_{\tau+1}$ 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

3

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 $d(t, t-1) = 1$ 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.

As a digression, it is worth noting that, with multiple breaks and a distance function that only takes on values 0 or 1, we are restricting the shift in coefficients to be of a similar magnitude each time a break occurs. This arises since σ_v is assumed to be constant over time. If this is objectionable, simple extensions are possible. For instance, in the two-break model, we can let $d(t, t-1) = 1$ if $t = \tau_1$ and $d(t, t-1) = \sigma_2$ if $t = \tau_2$ (where $\tau_2 > \tau_1$) and $d(t, t-1) = 0$ for all other values of t. Note that the introduction of the parameter σ_2 allows for the coefficient shift after the first break to be of a different magnitude than the coefficient shift after the second. Such extensions can be done for any of the multiple regime models we discuss below.

The structural break framework described in the previous paragraphs 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, Carter and Kohn (2000) and Giordani and Kohn (2006). As a simple example, if we let $d_t \equiv d(t, t-1)$ for $t = 1, ..., T$ be unknown parameters and use a hierarchical Bernoulli prior distribution for them:

$$
\begin{aligned}
p(d_t = 1) &= p, \\
p(d_t = 0) &= 1 - p
\end{aligned}
$$

where p is an unknown parameter, then we obtain the model of McCulloch and Tsay (1993). 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 d_t can be a vector (e.g. it can have two components, one controlling breaks in coefficients and the other in error variances). As long as the hierarchical prior for d_t has a Markov structure, the efficient algorithms of Gerlach, Carter and Kohn (2000) and Giordani and Kohn (2006) can be used. An important example of this would allow for $p(d_t)$ to have a form which restricts the usual matrix of Markov transition probabilities of Hamilton (1989). In this way, we can obtain the structural break variant of Hamilton's model developed in Chib (1998).

So far we have focussed on cases where $d(t, t-1)$ can take on only a few values (e.g. the values 0 or 1). However, by choosing other forms for the distance function, we can obtain intermediate values 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 were 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. With irregularly space data, the distance between observations matters in Harvey's derivations and our distance function plays the same role. This is a point we will develop further in the next section where we discuss TAR and smooth transition autoregressive (STAR) models.

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 (2006) and Koop and Potter (2007), there is a growing interest in structural break models lying between these extremes. Another related approach is 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 incorporate them all and let the data tell us which is preferred or do Bayesian model averaging over all the possibilities.

2.2   The Role of Data Reordering

By allowing for reordering of the data, we can accommodate various nonlinear time series models. It has long been recognized that many common nonlinear time series models are equivalent to models with structural breaks if the data is suitably reordered (e.g. Tsay, 1989). To illustrate how this is done, 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 $d\left(z_s^*, z_{s-1}^*\right) = 1$ if $z_{s-1}^* < \tau$ and $z_s^* \geq \tau$, and $d\left(z_s^*, z_{s-1}^*\right) = 0$ otherwise, then we obtain a two-regime TAR model (e.g. Potter, 1995):

$$y_t = \rho_1 y_{t-1} + \sigma_\varepsilon \varepsilon_t \text{ if } y_{t-1} < \tau$$
$$y_t = \rho_2 y_{t-1} + \sigma_\varepsilon \varepsilon_t \text{ if } y_{t-1} \geq \tau.$$

Multiple regime TAR models involve the obvious extension of this distance function definition. Standard approaches to TAR modeling assume a fixed number of regimes (e.g. two or three). However, just as with structural break models, it could be desirable to treat the number of regimes as unknown and estimate it. This can be done through hierarchical priors as discussed above. As a simple example, if we let $d_s \equiv d\left(z_s^*, z_{s-1}^*\right)$ and use a hierarchical Bernoulli prior distribution for these parameters:

$$p(d_s = 1) = p$$
$$p(d_s = 0) = 1 - p$$

we can obtain a TAR with an unknown number of regimes. Alternatively, various Markov switching models (see Hamilton, 1989) can be obtained (without re-ordering the data) by using a hierarchical prior which takes usual Markov transition probability form.

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. However, there is no reason for thinking why other index variables may not be possible. 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).

Another way of gaining useful interpretation for our approach is to relate it to nonparametric regression methods such as those described in Yatchew (1998). The intuition underlying such nonparametric approaches for cross-sectional regression is that, when approximating $y = f(x)$, if two observations have similar values for x then they have similar values for $f(x)$. If we use the distance functions $d(z_s^*, z_{s-1}^*) = 1$ for all s we obtain a model with similar intuition: since $\rho_s = \rho_{s-1} + \sigma_v v_s$, observations with similar values for z will have similar values ρ_s. Thus, if GDP growth is the dependent variable and if z is the lag of GDP growth, we have a model where periods with similar levels of lagged GDP growth will have similar dynamics. Or, equivalently, the dynamics of GDP growth gradually change with the business cycle. This example makes clear the usefulness of considering different index variables. That is, in addition to having a choice which implies "periods with similar levels of GDP growth *last quarter* will have similar AR dynamics", it could be empirically important to allow for a choice which implies "periods which have had similar levels of GDP growth *over the past year* will have similar AR dynamics" (i.e. $z_t = \frac{\sum_{j=1}^4 y_{t-j}}{4}$) or to allow for "periods where GDP growth *has accelerated rapidly* will have similar AR dynamics" (i.e. $z_t = \frac{\sum_{j=1}^p \Delta y_{t-j}}{p}$), or z could be the current-depth-of recession variable used in papers such as Beaudry and Koop (1993), etc. The possible choices for the index variable are endless and which ones are reasonable will depend on the empirical application at hand. But the important point to note is that the conventional literature simply selects a single choice for an index variable (or at most searches over a few choices by allowing for an unknown delay parameter). Our approach allows us to average over a wide range of possible choices or make the selection of index variable in a data-based manner without ex ante committing to a particular functional form.

As discussed in the preceding section, in the nonparametric regression literature, the distance between observations is also important, but this can easily be accommodated in our framework (e.g. by replacing $d(z_s^*, z_{s-1}^*) = 1$ by $d(z_s^*, z_{s-1}^*) = z_s^* - z_{s-1}^*$). In this regard, note also that the TAR models we have discussed so far all involve $d(z_s^*, z_{s-1}^*) \in \{0,1\}$ indicating when a regime-change has occurred. Allowing for $d(z_s^*, z_{s-1}^*)$ to take on any value in the unit interval yields the sort of gradual transition between regimes which is a characteristic

of STAR models. The relationship between our model and the STAR model is brought out empirically in our artificial data example below.

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 $d\left(z_s^*, z_{s-1}^*\right)$ to be treated nonparametrically, 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 reorderings of the data.[2] Accordingly, in this paper we restrict the set of allowable configurations for γ.

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] 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, Madigan, Raftery and Volinsky, 1999 or Chipman, George and McCulloch, 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.

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 $\tau_1, .., \tau_K$ $= 0$ 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[3]	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

2.3   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 $d\left(z_s^*, z_{s-1}^*\right) = z_s^* - z_{s-1}^*$. In this paper, we focus on the Normal distribution (with mean and variance estimated from the observed data using Bayesian

---

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

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 $d\left(z_s^*, z_{s-1}^*\right) = z_s^* - z_{s-1}^*$ 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 TAR model.[4] 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.

Now consider, for a fixed distance function, the flexibility that comes from the stochastic interpretation of the parameters. In the standard 2-regime TAR model the AR coefficients are constant within a regime. In the STAR model they evolve in a fixed way given by the parametric form of the transition function. Further, in both cases the posterior uncertainty about the nonlinear conditional expectation function is dependent on the parametric form. In our case we are not restricting the shape of the conditional expectation function. In addition, the posterior uncertainty about the nonlinear conditional expectation function responds to how dense the index observations are around the point. For example, consider a one step ahead prediction at a point z in the interval $\left(z_s^*, z_{s-1}^*\right)$ which depends crucially on $\rho_{g(z)}$. It can be shown that the mean of $\rho_{g(z)}$ interpolates between the two neighboring values of ρ with weights inversely related to the distance between neighboring points and z. The variance shares similar properties and can be large if the "distance" between observations is large. In general, arguments such as this show how our approach is almost nonparametric in spirit, providing much more flexible and robust predictions than standard parametric approaches.

Another advantage of our modeling framework is that it is simple, being based on the standard state space model. Thus, textbook methods for estimation, model comparison and prediction are available. In this paper, we apply Bayesian methods using a Markov Chain Monte Carlo (MCMC) algorithm for posterior simulation. Conditional on a particular ordering of the data (i.e. conditional on γ) and a particular distance function, posterior simulation is easy, involving commonly-known Bayesian simulation methods for state space models (e.g. Durbin and Koopman, 2002 or DeJong and Shephard, 1995) and simulation methods for stochastic volatility (e.g. Kim, Shephard and Chib, 1998). Thus, we only require a method for drawing from the posterior for γ and the parameters of the distance function (conditional on the other parameters in the model). This method is supplied in the following section.

This completes our discussion of the basic intuition behind our model framework and the general ideas underlying our posterior simulator. To summarize: we take a standard state space model and then argue that, by allowing for different orderings of the data and allowing for variances in the state equation that depend upon the distance between observations, we

---

[4]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.

have created an extremely flexible model which nests a huge variety of popular nonlinear time series and structural break specifications (as well as many other time series models). Methods for Bayesian inference are very simple, involving only minor additions to standard Bayesian methods for state space models.

In the next section, we provide formal details to make rigorous the informal intuition provided in this introduction. The model considered is a generalization of the one discussed so far, freeing up the assumption that x_t is a scalar. Furthermore, we allow for the conditional variance of the measurement equation to depend on γ, thus extending the concept of stochastic volatility to allow for more general nonlinear patterns in the conditional variance. Many recent empirical macroeconomic studies (e.g. Cogley and Sargent, 2005 and Sims and Zha, 2006) have emphasized the statistical importance (and relevance for macroeconomic policy) of allowing for stochastic volatility or nonlinearities in the conditional variance. We develop methods for drawing from γ and the parameters of the distance function (conditional on the other parameters of the model) and discuss various ways of restricting the set of configurations γ can take on. We apply our methods to artificial data and in a study of the dynamics of real GDP growth.

3   Bayesian Inference

Consider the 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 row 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 $\theta = (\theta_0', .., \theta_T')'$ (conditional

on $\alpha = (\alpha_0, .., \alpha_T)'$ and Q).[5] The method of Kim, Shephard and Chib (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, Shephard and Chib (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 $p\left(\theta | Data, \alpha, Q, \sigma_\eta^2\right)$, $p\left(\alpha | Data, \theta, Q, \sigma_\eta^2\right)$, $p\left(Q^{-1} | Data, \theta, \alpha, \sigma_\eta^2\right)$ and $p\left(\sigma_\eta^{-2} | Data, \theta, \alpha, Q\right)$.[6] The results of Fernandez, Ley and Steel (1997) imply that proper priors for either of the error variances or the initial conditions in state equations are required in order to obtain proper posteriors.[7] 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 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

$$y_t = X_t' \theta_{g(z_t)} + \varepsilon_t, \tag{12}$$

where

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

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

$$\varepsilon_t = \xi_t \exp\left(\frac{1}{2} \alpha_{g(z_t)}\right) \tag{14}$$

where

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

---

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

[6]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 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 $p\left(Q^{-1} | Data, \theta, \alpha\right)$.

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

and $\eta_s \sim N\left[0, \sqrt{d\left(z_s^*, z_{s-1}^*\right)}\sigma_\eta^2\right]$. Additional assumptions about the errors are as described above (see equations 8 through 11 and surrounding discussion). Other definitions are as after (4). Briefly s indexes a reordering of time according to $g\left(z_t\right)$ which is the rank of the index variable, z_t, when $z_1, ..z_T$ are placed in ascending order. The reordered values of z_t are denoted as z_t^* (for $t = 1, ...T$) and, by definition, we have $z_1^* \le z_2^* \le .. \le z_T^*$. We introduce a parameter vector γ to denote how the index variable orders the data, the tth element of γ is $\gamma_t = g\left(z_t\right)$ for $t = 1, ..., T$. $d\left(z_s^*, z_{s-1}^*\right)$ 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 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 orderings for both the conditional mean and conditional variance, the computational demands would increase greatly. As we shall explain in the empirical section, in most cases one could focus on the case where the conditional variance is only based on the natural time ordering.

We use a distance function based on Normal CDFs (see equation 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 $p\left(\theta|Data, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2, \gamma\right)$, $p\left(\alpha|Data, \theta, Q, \sigma_\eta^2, \mu, \sigma_d^2, \gamma\right)$, $p\left(Q^{-1}|Data, \theta, \alpha, \sigma_\eta^2, \mu, \sigma_d^2, \gamma\right)$, $p\left(\sigma_\eta^{-2}|Data, \theta, \alpha, \sigma_\eta^2, \mu, \sigma_d^2, \gamma\right)$, $p\left(\mu|Data, \theta, \alpha, Q, \sigma_\eta^2, \sigma_d^2, \gamma\right)$, $p\left(\sigma_d^2|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \gamma\right)$ and $p\left(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2\right)$. The first four of these posterior conditional distributions are standard (see the discussion surrounding the model defined by equations 8 through 11). Although note that some minor modifications of the algorithm of Kim, Shephard and Chib (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 $\sqrt{d\left(z_s^*, z_{s-1}^*\right)}$ we obtain a state equation for volatilities that is in the same format as Kim, Shephard and Chib (1998). The final three of these posterior conditionals we discuss here.

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

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

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 $p\left(\theta, \alpha|\mu, \gamma\right)$, which by (13) and (15) is Normal, and the prior, $p\left(\mu\right)$. 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:

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

where $(s-1)$ superscripts (such as $\mu^{(s-1)}$) denote the $(s-1)^{th}$ 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:

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

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. This is of the same form as the algorithm for μ except that (17), (18) and (19) are replaced with formula involving σ_d^2 instead of μ. To be precise, using similar steps as in (17) we have:

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

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 Random Walk Chain Metropolis-Hastings algorithm proceeds analogously to (18) and (19), the only difference is that we replace (18) by:

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

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 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 $p(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2)$). 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 $p(\gamma|Data, \theta, \alpha, Q, \sigma_\eta^2, \mu, \sigma_d^2)$). 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 the approach of Gelfand and Dey (1994).

In our empirical work with a TVP extension of an AR(p) model, we restrict the possible configurations of γ:

- $\gamma = (1, 2, 3, .., T)'$, the real time ordering (i.e. $z_t = t$)

- data ordered according to the jth lag of the dependent variable or covariates (i.e. $z_t = y_{t-j}$) for $j = 1, .., p$.

- data ordered to average of the lagged dependent variable or covariates over the last p periods (i.e. $z_t = \frac{\sum_{j=1}^{p} y_{t-j}}{p}$)

Thus, the number of possible configurations in a standard univariate time series model is $2p$. Of course, in another empirical exercise other sets of configurations are 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 model are familiar ones (e.g. AR coefficients, error variances) 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 $z_t = z_t^* = t$ for $t = 1, ..., T$. 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 $1.$[8] 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 Figure 1 presents the prior mean of the distance function for every value of z_t^*. 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 Figure 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 Figure 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 Figure 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 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 $t = 1, .., 200$ where ε_t is i.i.d. $N(0, 1)$. The coefficients evolve according to:

---

[8]We find that some form of standardization produces more efficiency in our MCMC algorithm. Equation 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 $z^* = -3$, a maximum value of $z^* = +3$ and for example a divisor of $2T$.

$$F\left(z_{t}\right)=\frac{1}{1+\exp\left(-az_{t}\right)}. \tag{23}$$

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 $\sigma_\varepsilon = .01$. 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 reordering the data at this stage. Thus our first data set involves the choices above plus $z_t = t$ for $t = 1, ..., 200$. 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 equations 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 $W\left(\nu, H\right)$ 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 equation 21). The priors for μ and σ_d^2 were described in the previous sub-section.

Using the methods of posterior simulation described above, with $X_t = y_{t-1}$ and index definition variable simply being the natural ordering (i.e. $1, 2, .., T$), we can obtain posterior properties of any of the model parameters (or functions thereof). Figures 2 and 3 graph some aspects of particular interest. Figure 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. Figure 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 reordering, 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 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.

Figure 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 no reordering of the data and the distance function defined to always be one. We use the same prior for the TVP parameters as for the parameters of our model. Figure 5 is comparable to Figure 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 BIC's 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 data reordered by z_t, we estimate the distance function using this ordering. Since our approach involves multiple choices for z_t, Figure 6 transforms back to the natural time ordering. This accounts for the irregular shape of Figure 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.

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, 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, Nelson and Piger (2003) 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.

The best fitting linear model is the AR(2) and, hence, we consider extensions of this model and let X_t contain an intercept plus 2 lags of y_t. 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 $\theta_0 \sim$
$N\left(0_3, \begin{bmatrix} 10^{10} & 0 & 0 \\ 0 & 0.7^2 & 0 \\ 0 & 0 & 0.7^2 \end{bmatrix}\right)$ which still reflects a (weak) prior belief in the stationarity.
For the stochastic volatility component we assume σ_η^{-2} to be $G(10, 50)$.

With our model we find strong evidence in favour 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. Figure 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(2) and, hence, we will not present results for an AR(2). 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) through (11) or, equivalently, our model with $\gamma = (1, ..., T)'$ and $d(z_s^*, z_{s-1}^*) = 1$ for all observations. Given the fact that our model yields strong evidence in favour of $\gamma = (1, ..., T)'$, it is not surprising that we find results to be similar between our approach and the TVP model. Figure 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 Figures 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 0.787 for our model, but 0.681 for the TVP model (and only 0.342 for the 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:

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

before presenting convincing evidence of departures using a flexible approach to nonlinear inference developed in Hamilton (2001). Note that this approach focusses on nonlinearities in the conditional mean (with the conditional variance being of less interest). We can view this equation as one of an unmodeled larger VAR system (possibly with restrictions). Thus, we label the conditional mean estimates other than the intercept, VAR coefficients.

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.[9] Our set of explanatory variables (i.e. X_t in equation 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 conditional mean. 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),

---

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

we find overwhelming evidence in favor of $\gamma = (1, ..., T)'$. In light of this (and in order to illustrate an empirically useful extension), in this section we present results where different orderings and distance functions exist for the conditional mean (i.e. the VAR coefficients) and the conditional variance (i.e. the state equations 13 and 15 can be based on different orderings). That is, we now have a γ_1 (which controls ordering relating to the conditional mean) and γ_2 (for the conditional variance). 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 $\gamma_2 = (1, ..., T)'$ is so predominant, the results below just use this choice. For our modified state space model defined in (12) through (15), we now have two distance functions so that $v_s \sim N \left[0, \sqrt{d_1 \left(z_s^*, z_{s-1}^* \right)} Q \right]$ and $\eta_s \sim N \left[0, \sqrt{d_2 \left(z_s^*, z_{s-1}^* \right)} \sigma_\eta^2 \right]$.

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 for the conditional mean (i.e. different choices for γ_1). In contrast to our univariate results, there is now a certain degree of uncertainty over data orderings (where we use the marginal likelihood for each ordering to obtain a probability for the ordering). Roughly 5% of the probability is associated with the choice which implies the VAR coefficients evolve according to a conventional TVP model. However, most of the probability is attached to models which order observations according to past oil price inflation. The most probable ordering is based on the index variable $z_t = \frac{\sum_{j=1}^{4} y_{t-j}}{4}$. 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. For the non-Bayesian, Table 2 also presents the log-likelihoods evaluated at the maximum likelihood estimate (MLE) obtained using each index variable. These show a similar pattern to the Bayesian probabilities. If we were to treat each choice for γ_2 as defining a model, then any conventional non-Bayesian hypothesis testing/model selection procedure would choose the model based on the index variable $z_t = \frac{\sum_{j=1}^{4} y_{t-j}}{4}$.

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 . As a further illustration of this, if we had included the TVP model in our set of models given in Table 2, its marginal likelihood (using the same prior as for our model for the parameters they share in common) implies that it would have received only 2.3% of the probability.

We remind the reader that, with regards to the measurement equation error variance, we are assuming a stochastic volatility specification and simply setting $\gamma_2 = (1, ..., T)'$ (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 Figure 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.

Table 2: Probability of Each Ordering and Posterior Properties of Key Parameters (numbers in parentheses are posterior standard deviations)						
Index for Cond. Mean			Dist. Function Cond. Mean		Dist. Function Cond. Variance	
	Prob.	Log Like	$E(\mu)$	$E(\sigma_d^2)$	$E(\mu)$	$E(\sigma_d^2)$
t	0.052	-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)
$\frac{\sum_{j=1}^{3} y_{t-j}}{3}$	0.021	-49.46	0.180 (0.424)	1.227 (0.487)	0.277 (0.586)	5.458 (4.266)
$\frac{\sum_{j=1}^{4} y_{t-j}}{4}$	0.005	-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}	0.001	-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)
$\frac{\sum_{j=1}^{3} p_{t-j}}{3}$	0.097	-45.26	0.857 (0.409)	2.406 (1.312)	-0.081 (0.571)	4.470 (4.720)
$\frac{\sum_{j=1}^{4} p_{t-j}}{4}$	0.821	-43.73	0.914 (0.561)	2.177 (0.857)	-0.282 (0.638)	4.268 (4.667)

Thus, far we have said little of the economic implications of our findings. Figure 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. $\beta_5 + .. + \beta_8$). 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. Figure 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.[10] 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 Figure 10, 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, Figure 11 is the same as Figure 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. Note that Figure 11 is more consistent with Hamilton's story that positive oil shocks have larger marginal effects on output than do negative oil shocks.

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 and/or the conditional variance). 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 ordering and distance. By ordering the data in various ways, 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 involve, 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

---

[10]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 Figure 10.

literature. Our work with artificial and real data show the advantages of our approach.

6 References

Bai, J. and Perron, P. (1998). "Estimating and testing linear models with multiple structural changes," *Econometrica* 66, 47-78.

Beaudry, P. and Koop, G. (1993). "Do recessions permanently change output?" *Journal of Monetary Economics*, 31, 149-164.

Blanchard, O. and Simon, J. (2001). "The long and large decline in US output volatility," *Brookings Papers on Economic Activity*, 1, 135-174.

Carlin, B. and Chib, S. (1995). "Bayesian model choice via Markov Chain Monte Carlo methods," *Journal of the Royal Statistical Society, Series B*, 57, 473-484.

Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, second edition. Boca Raton: Chapman and Hall.

Chib, S. (1998). "Estimation and comparison of multiple change-point models," *Journal of Econometrics,* 86, 221-241.

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327-335.

Chib, S. and Jeliazkov, I. (2001). "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270-281.

Chipman, H., George, E. and 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. and Sargent, T. (2001). "Evolving post World War II inflation dynamics," *NBER Macroeconomics Annual*, 16, 331-373.

Cogley, T. and Sargent, T. (2005). "The conquest of US inflation: Learning and robustness to model uncertainty," *Review of Economic Dynamics*, 8, 528-563.

DeJong, P. and Shephard, N. (1995). "The simulation smoother for time series models," *Biometrika*, 82, 339-350.

van Dijk, D., Giordani, P. and Kohn, R. (2007) "A unified approach to nonlinearity, outliers and structural breaks," *Journal of Econometrics,* 137, 112-133.

Durbin, J. and Koopman, S. (2002). "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603-616.

Fernandez, C., Osiewalski, J. and 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. and 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. and Kohn, E. (2000). "Efficient Bayesian inference in dynamic mixture models," *Journal of the American Statistical Association*, 95, 819-828.

Giordani, P. and Kohn, R. (2006). "Efficient Bayesian inference for multiple change-point and mixture innovation models," manuscript.

Green, P. (1995). "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711-732.

Hamilton, J. (1989). "A new approach to the economic analysis of non stationary time series and the business cycle," *Econometrica*, 57, 357-384.

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. and Koopman, S. (2000). "Signal extraction and the formulation of unobserved components models," *Econometrics Journal*, 3, 84-107.

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382-417.

Kim, C., Nelson, C. and Piger. J. (2003). "The less volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations," working paper 2001-016C, The Federal Reserve Bank of St. Louis.

Kim, S., Shephard, N. and 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. and Potter, S. (1999). " Dynamic asymmetries in US unemployment," *Journal of Business and Economic Statistics*, 17, 298-312.

Koop, G. and Potter, S. (2001). "Are apparent findings of nonlinearity due to structural instability in economic time series?" *The Econometrics Journal*, 4, 37-55.

Koop, G. and Potter, S. (2006). "The vector floor and ceiling model," Chapter 4 in *Nonlinear Time Series Analysis of the Business Cycle*, edited by C. Milas, P. Rothman and D. van Dijk in Elsevier's *Contributions to Economic Analysis* series.

Koop, G. and Potter, S. (2007). "Estimation and forecasting in models with multiple breaks," *Review of Economic Studies*, forthcoming.

Lundbergh, S., Terasvirta, T. and van Dijk, D. (2003). "Time-varying smooth transition autoregressive models," *Journal of Business and Economic Statistics*, 21, 104-121.

McConnell, M. and Perez, G. (2000). "Output fluctuations in the United States: What has changed since the early 1980s?" *American Economic Review,* 90, 1464-76.

McCulloch, R. and 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.

Rahbek, A. and Shephard, N. (2002). "Inference and ergodicity in the autoregressive conditional root model," manuscript available at http://www.nuff.ox.ac.uk/users/shephard/.

Sims, C. and Zha, T. (2006). "Were there regime switches in macroeconomic policy?" *American Economic Review*, 96, 54-81.

Stock J. and Watson M. (2002). "Has the business cycle changed and why?" *NBER Macroeconomic Annual.*

Stock J. and Watson M. (2007). "Is inflation harder to forecast?" *Journal of Money, Credit and Banking*, forthcoming.

Tierney, L. and 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.

Yatchew, A. (1998). "Nonparametric regression techniques in economics," *Journal of Economic Literature*, 36, 669-721.
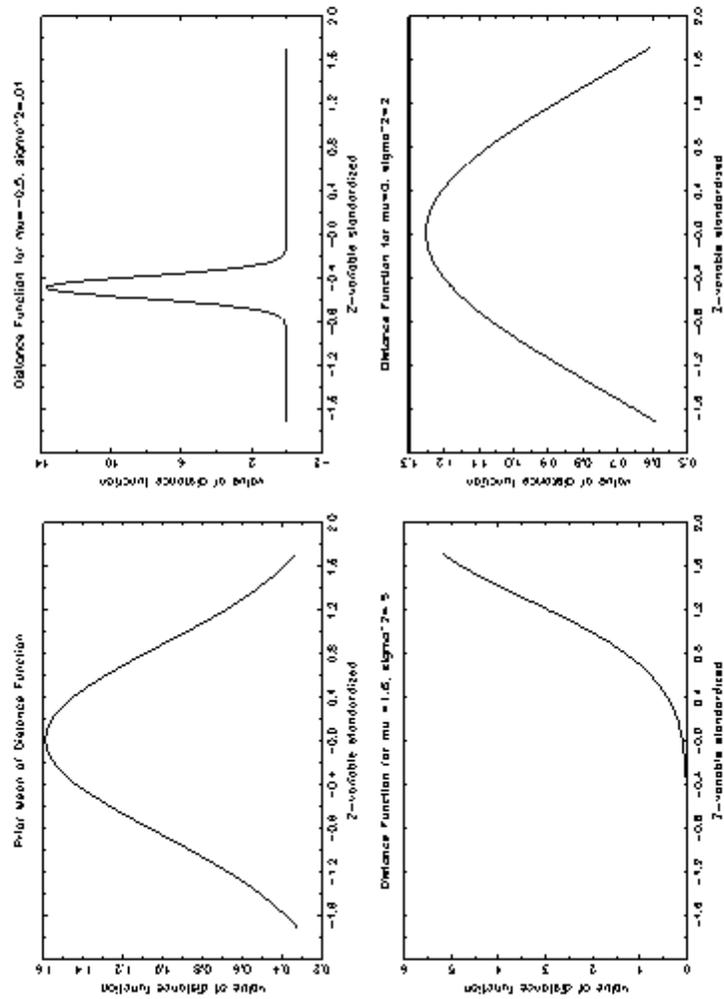
Figure 1: Prior Properties of Distance Function
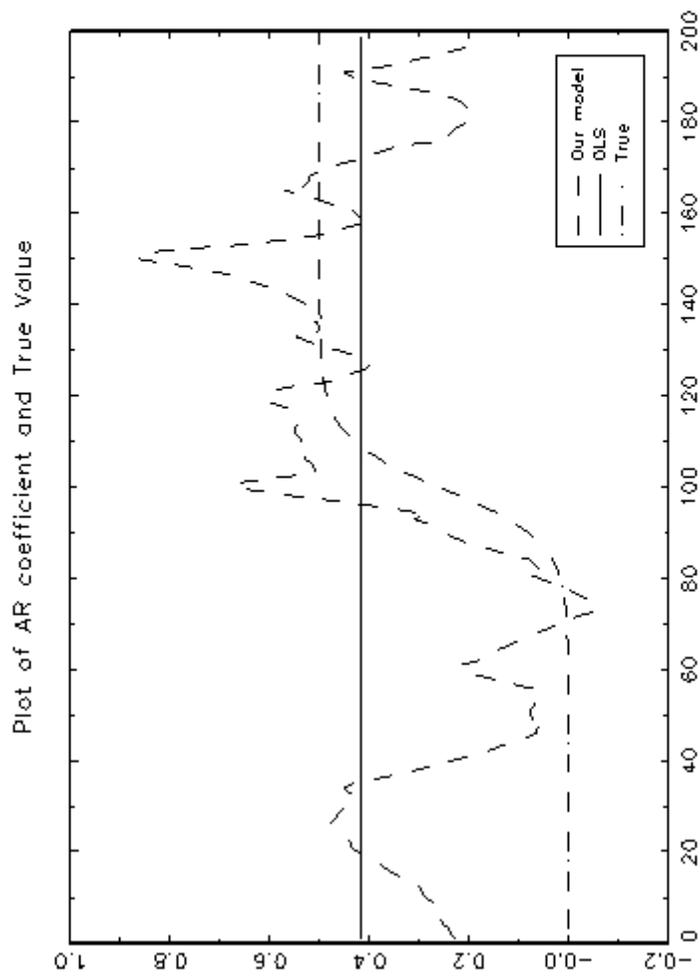
Figure 2: The AR Coefficient for Data Set 1
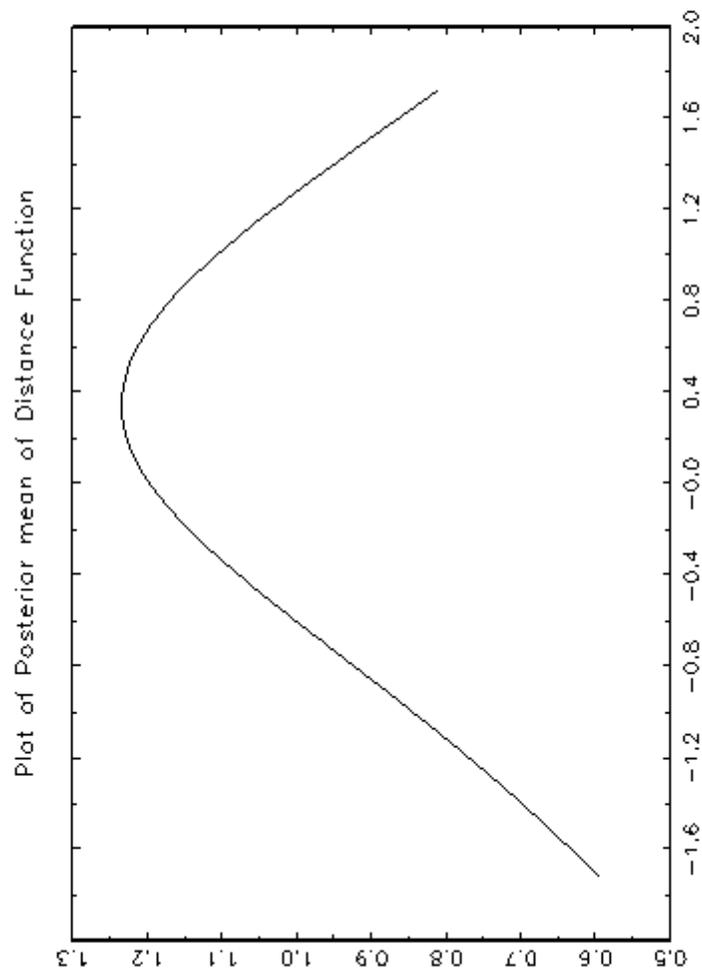
Figure 3: The Distance Function for Data Set 1
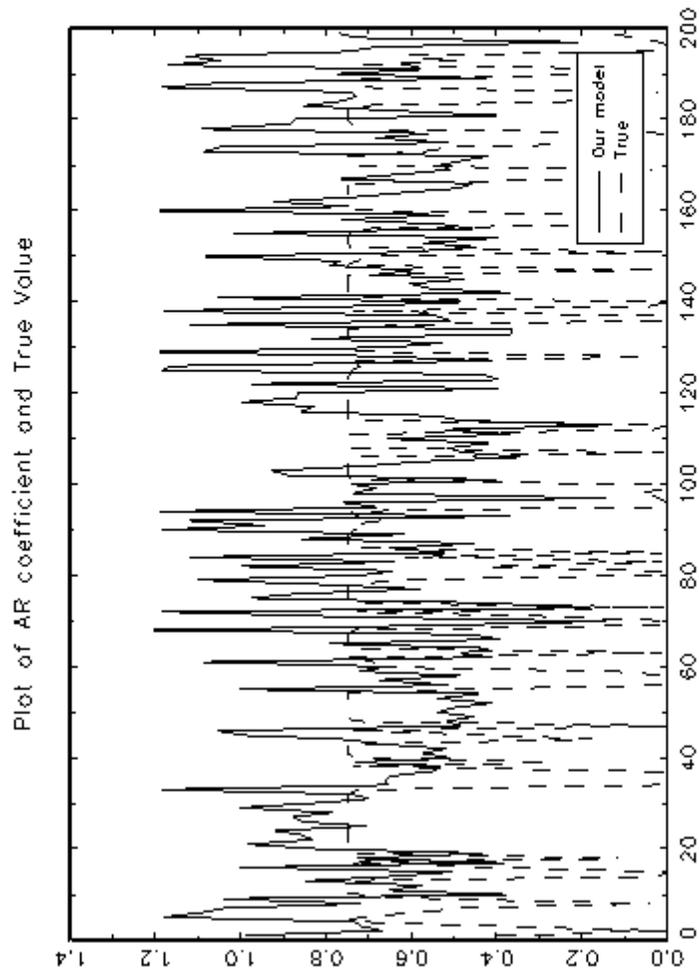
Figure 4: The AR Coefficient for Data Set 2 (Our Model)
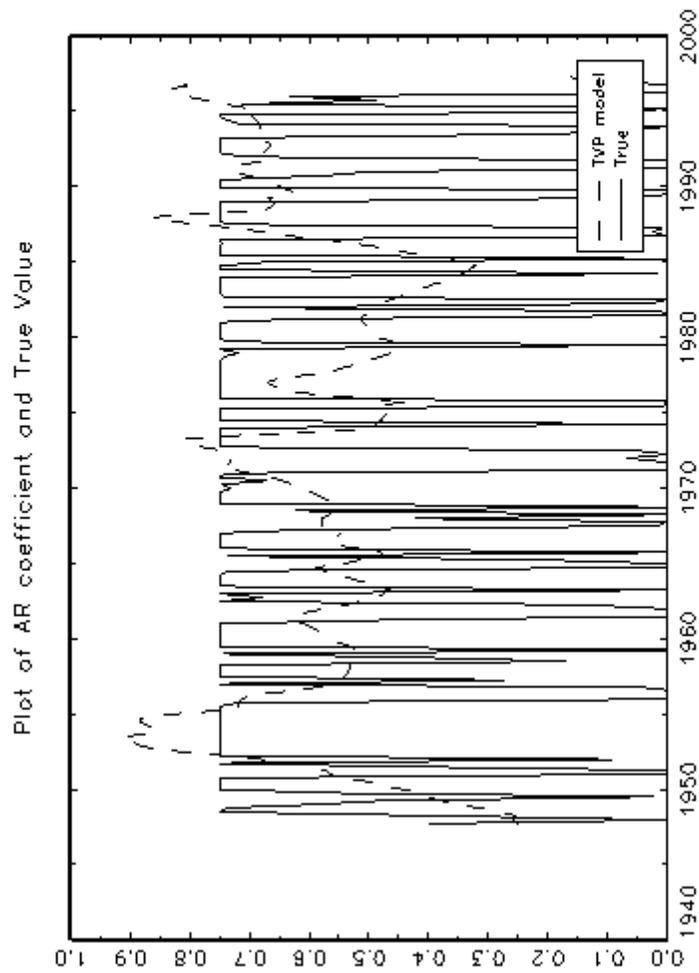
Figure 5: The AR Coefficient for Data Set 2 (TVP Model)
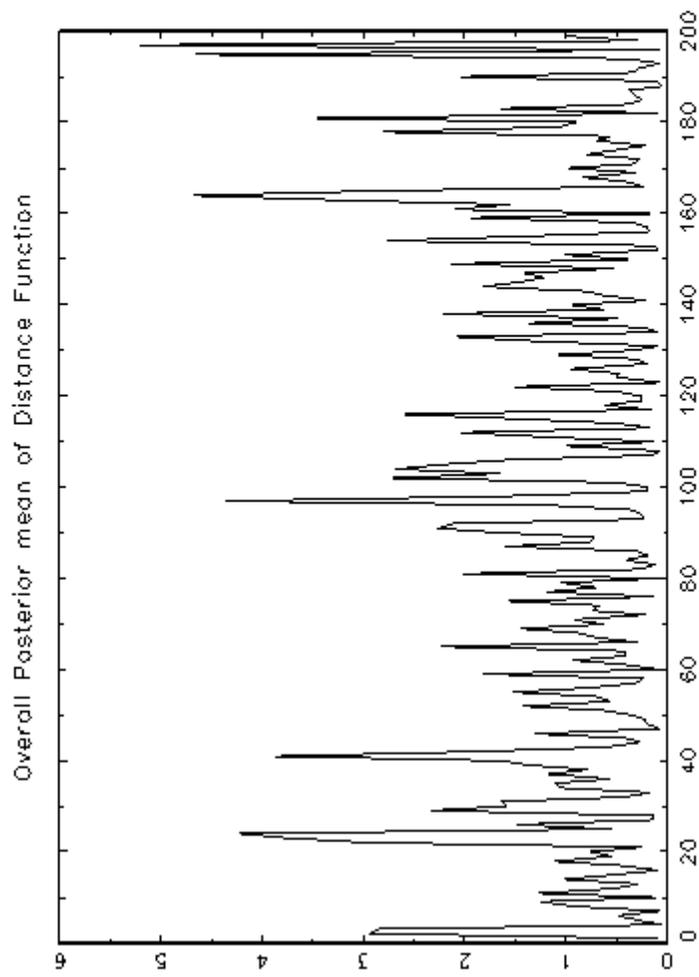
Figure 6: The Distance Function for Data Set 2
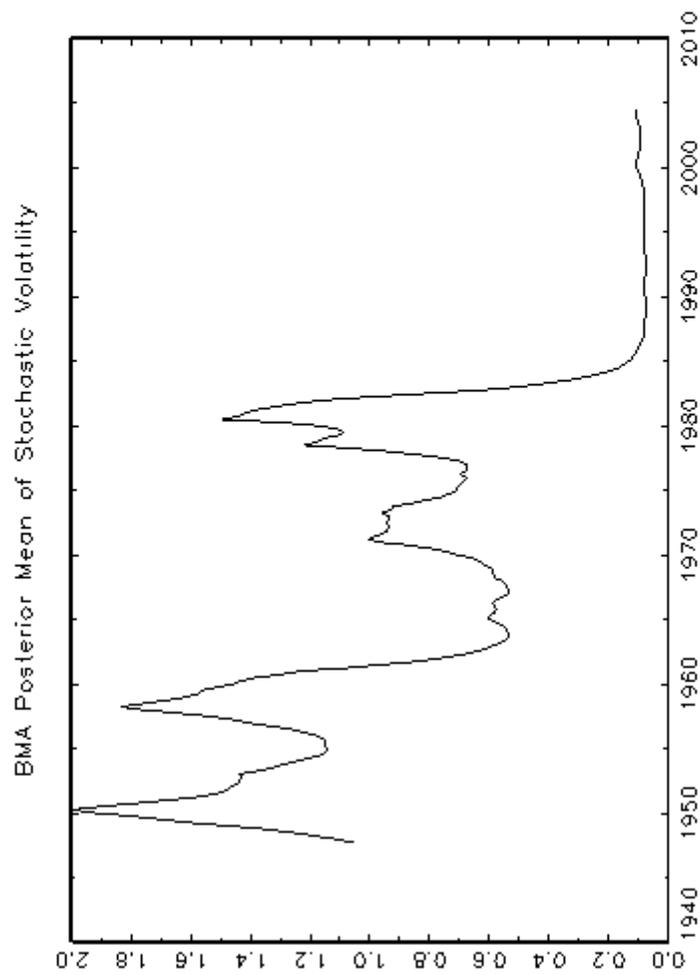
Figure 7: Stochastic Volatility (univariate example)

33
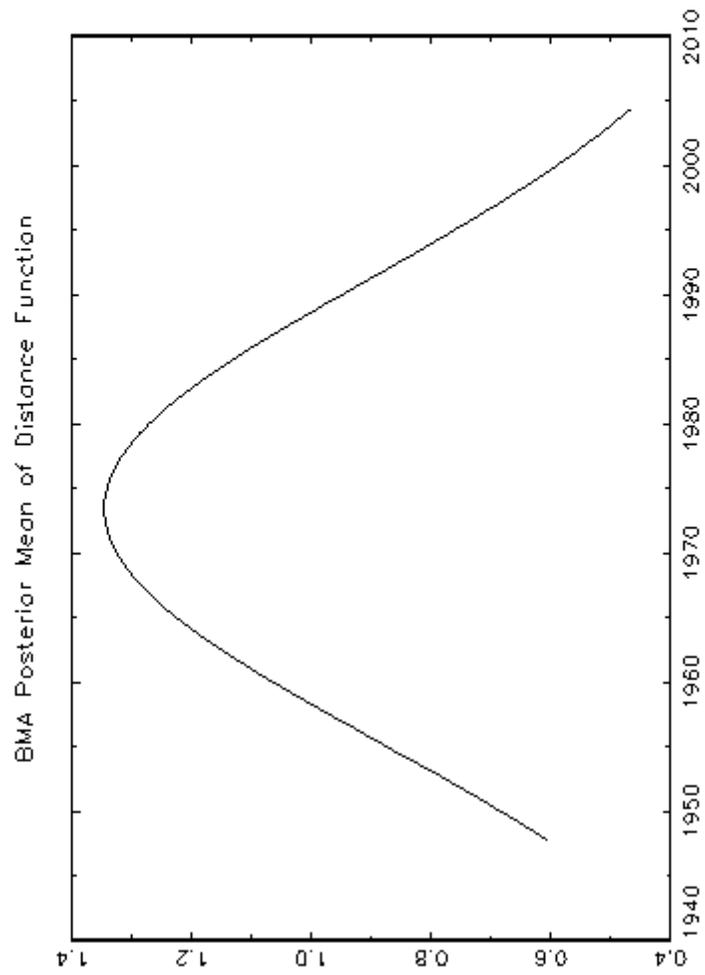
Figure 8: Distance Function (univariate example)

34
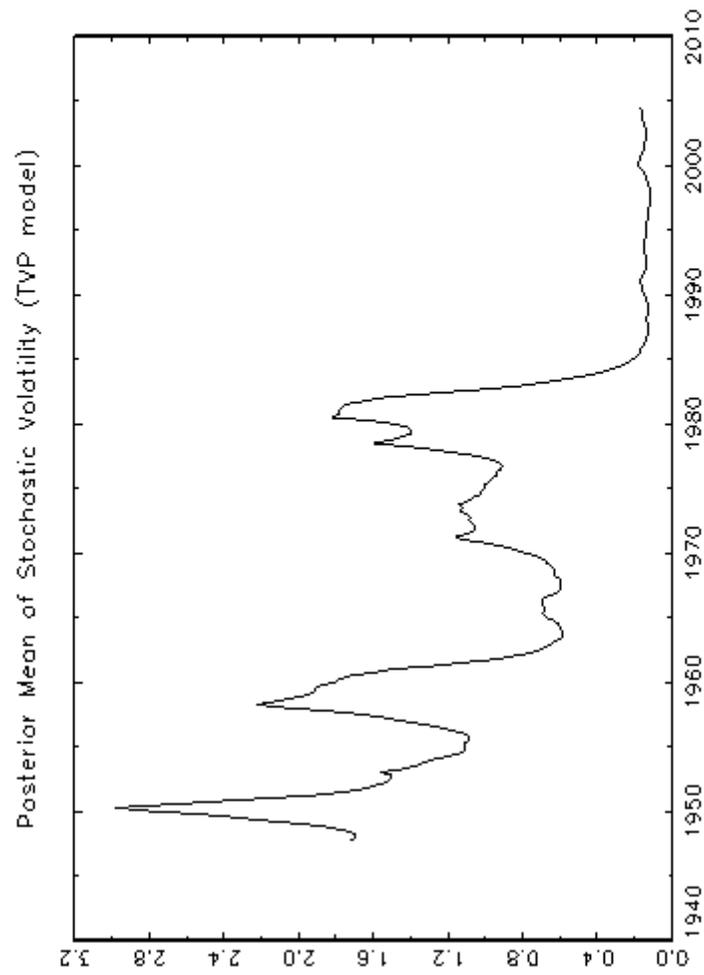
Figure 9: Stochastic Volatility for TVP Model (univariate example)
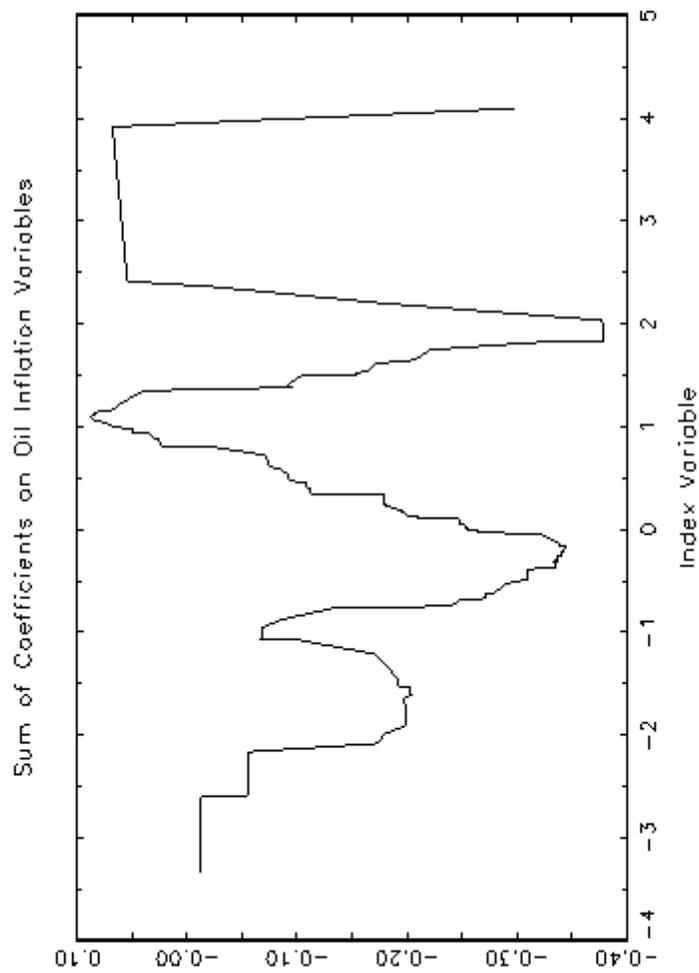
Figure 10: A Measure of the Effect of an Oil Shock for the Preferred Ordering
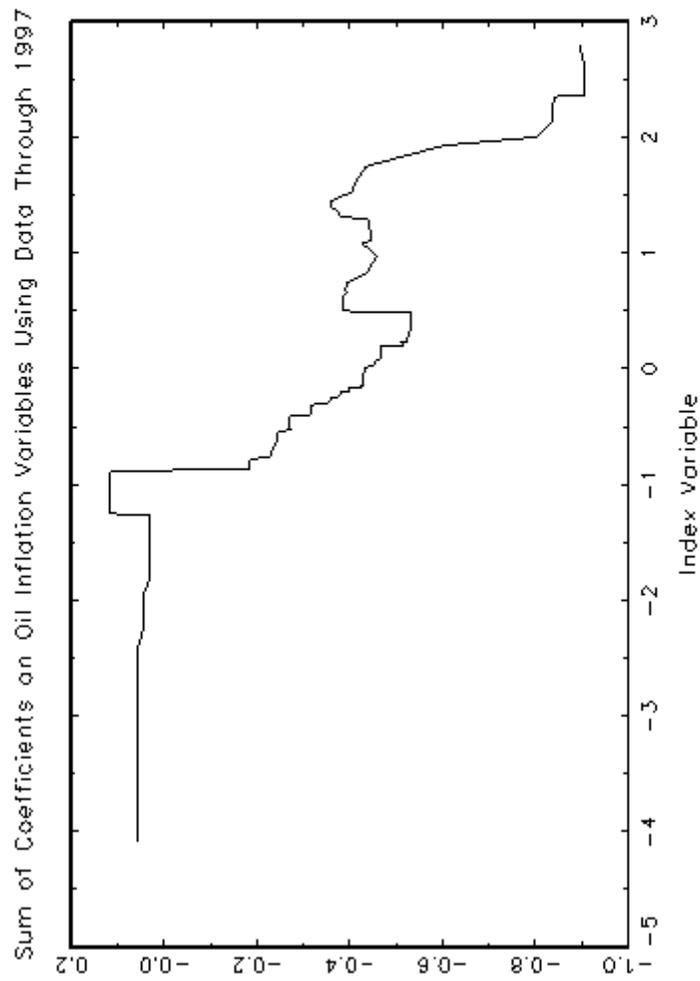
Figure 11: A Measure of the Effect of an Oil Shock for the Preferred Ordering Using Data Through 1997

37