NONLINEARITY, STRUCTURAL BREAKS OR OUTLIERS IN ECONOMIC TIME SERIES?*

GARY KOOP †
Department of Economics, University of Edinburgh,
50 George Square
Edinburgh, EH8 9JY, U.K.
E-mail: G.Koop@ed.ac.uk

SIMON POTTER ‡
Department of Economics, UCLA,
405 Hilgard Avenue
Los Angeles, CA 90095-1477, U.S.A.
E-mail: potter@ovett.sscnet.ucla.edu

August 1997

*This paper has its motivation from discussions at the EC^2 conference in 1995.

†Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged, as are helpful discussions with Sid Chib and Les Oxley.

‡Financial support from the Academic Senate UCLA and Center for Computable Economics, UCLA is gratefully acknowledged.

In recent years there has been an increasing interest in nonlinear models as an alternative to the linear specifications which have dominated the applied macroeconomics literature. For many series empirical evidence for nonlinearity exists. However, it is possible that this apparent nonlinearity could be due to structural breaks or outliers. Hence, this paper develops methods for comparing linear, nonlinear, structural break and outlier models. We adopt a Bayesian approach which allows for the easy comparison of non-nested models and surmounts the problems associated with nuisance parameters which are unidentified under the null which plague classical tests. The computational difficulties associated with the Bayesian approach are surmounted by working with autoregressive switching models for which analytical posterior results for most of the parameters are available. After motivating and deriving Bayesian methods for such models, an empirical section analyzes the behaviour of the growth of US real GDP and British industrial production.

Keywords: Bayesian, Marginal Likelihood, threshold autoregressive model, changepoint problem, structural instability, outliers

1 Introduction

Time series models are often used to predict or improve understanding of the dynamic properties of macroeconomic data. Parametric models are commonly used which assume a constant linear dynamic structure over time. For instance, much of modern macroeconomics, including the unit root literature, have been based on autoregressive or autoregressive moving average models. However, this assumption may be inappropriate. For example, U.S. post-World War II macroeconomic series cover a time span which includes wars, rapid technical change, the breakdown of the Bretton Woods agreement, the change in the Fed's operating behavior under Paul Volcker and numerous other shifts in monetary and fiscal policy.

All of these factors may imply that it is inappropriate to assume the same dynamic model exists for the 1950's, say, as for the 1980's. We refer to models where the dynamics change permanently in a way that cannot be predicted by the history of the series as "structural break" models. Alternatively, it is possible that dynamic properties can vary over the business cycle. For instance, it is likely that shocks to real output have different effects if the economy is operating below capacity (i.e. is in a recession), relative to cases where the economy is in normal times or is overheating (i.e. is expanding rapidly). With some abuse of terminology, we refer to models which allow for dynamics which vary over the business cycle in a predictable way as "nonlinear" models. Still another possibility is that apparent departures from linearity are due to unpredictable large shocks which have only temporary effects. We refer to models which have this property as "outlier" models.

Nonlinear models provide very different understanding of the effects of shocks over the business cycle and lead to forecasts depending on the state of economic activity. For example, as shown by Beaudry and Koop (1993), a simple nonlinear model implies that positive shocks to U.S. output growth are more persistent than negative ones. In contrast, structural break models have no predictable changes in regime over the business cycle.

Hence, a consequence of adopting them is that the current forecasting model is estimated only using data observed since the most recent break. Outlier models are useful for removing the influence of rare events from the estimates used for forecasting and analyzing dynamics.

Since they have very different consequences for forecasting and understanding business cycles, it is important to test whether nonlinearities, structural breaks or outliers are present in economic time series, and if they are, to estimate models which incorporate them. A significant literature now exists which tests for structural breaks or nonlinearity in macroeconomic time series. However, this literature, with a few exceptions, takes a classical econometric approach and concentrates on only one of the three possible classes of model considered in this paper. For example, an important recent paper, Stock and Watson (1996), uses a battery of classical tests on 76 US postwar quarterly time series and finds significant evidence of structural instability in many cases. However, it is possible that this apparent structural instability is, in reality, a reflection of some form of nonlinearity. Hence, it is important not only to compare linear to structural break models, but also to compare structural break to nonlinear models and outlier models.

Such a comparison of many possibly non-nested models is theoretically challenging using classical techniques but theoretically straightforward using Bayesian approaches. Bayesian methods for choosing between models are based on the marginal likelihood associated with each model. The marginal likelihood is the average of the likelihood function across parameters values with respect to the prior distribution of the parameters. The comparison of multiple, possibly non-nested models, can be accomplished by constructing posterior model probabilities directly from these marginal likelihoods.[4] The number or nature of the models under consideration does not affect the logic of the calculation. The two main drawbacks to such a Bayesian approach are the need to specify informative priors and the

[4]That is, under the assumption of equal prior probabilities for each model, one adds the marginal likelihood across all models under consideration and then divides the individual marginal likelihoods by the value of this sum.

difficulty in calculating the marginal likelihood. Before discussing these drawbacks it is useful to outline other advantages of the Bayesian approach for nonlinear, structural break and outlier models.

1.1 Advantages of a Bayesian Approach

In previous work (Koop and Potter, 1996) we note that, with many nonlinear models, likelihood functions are non-smooth and multimodal. Similar observations apply to structural break and outlier models. Bayesian methods, by using information from the entire parameter space capture this finite sample uncertainty about the true parameter values. In contrast, standard classical maximum likelihood methods choose one point (i.e, the MLE in sample) in the parameter space and use a normal asymptotic approximation to capture the local uncertainty around this point. Uncertainty produced by multiple peaks in the likelihood function is ignored.

In addition, posterior model probabilities can be used to combine dynamic features of different models. For example, instead of 'accepting' or 'rejecting' the linear model one can include its dynamics weighted by the posterior probability for linearity. This means that forecasts and impulse response functions will be more reflective of the underlying model and parameter uncertainty when based on Bayesian methods of analysis.

Another issue is the choice of lag length. It is standard to choose a particular lag length at which to conduct the analysis by using information criteria or by testing the significance of additional lags. This leads to issues of data-mining and reduces the credibility of the results. Rather than picking a particular lag length to work with Bayesian methods allow one to combine information from a range of lag lengths for each model.

In constructing classical statistical tests of the null hypothesis of a linear model versus the three alternatives one runs into Davies' problem: nuisance parameters which are not identified under the null. This leads to the difficult classical statistical issues addressed in Andrews (1993) and Andrews and Ploberger (1994). Optimal classical solutions to Davies' problem involve integrating out nuisance parameters with respect to an *a priori* weighting function and calculation of critical values for classical test statistics can be computationally demanding – involving complicated simulation methods (see Hansen, 1996). Koop and Potter (1996) shows how the presence of nuisance parameters that are unidentified under the null poses no problems for a Bayesian analysis and the unidentified nuisance parameters can be integrated out using an *a posteriori* weighting function.

In the case of structural break and/or outlier models Bayesian methods, unlike the classical approach, extend directly to making inferences regarding the possibility of multiple breaks or outliers.

1.2 Drawbacks of the Bayesian Approach

A major drawback of using Bayesian methods is that they can be very computationally demanding. For instance, in the case of structural break models, they require specification of a parametric likelihood function before and after the breakpoints (see, eg., Barry and Hartigan, 1993, Carlin, Gelfand and Smith, 1992 and Stephens, 1994). For many common specifications (e.g. ARMA models, regression models with AR errors, models with non-Normal errors, etc.), analytic posterior properties do not exist and Bayesian analysis requires the use of simulation-based methods such as the Gibbs sampler. This implies, if the breakpoint is unknown, that likelihood evaluations at every possible breakpoint are often required at every pass through the Gibbs sampler. If the breakpoint can occur at any time, the sample size is T, and the Gibbs sampler is run for S passes, then approximately ST likelihood evaluations must be made. For typical values of S and T (eg. 10,000 and 200), the computational burden becomes quite large. If two breakpoints are allowed for, the number of likelihood function evaluations becomes approximately ST^2. In addition, researchers are often interested in investigating many different models, further increasing the computational burden.

In the present paper, for one of our data sets, we work with 6 different lag lengths and allow for 0, 1 and 2 breakpoints. Furthermore, we consider nonlinear and outlier models with homoscedastic and heteroscedastic versions of most models. In total we estimate 66 models. In cases of this sort, any Bayesian procedure requiring extensive simulation is virtually impossible given the present level of computer technology. Similar arguments hold for many common nonlinear specifications (eg. the Markov switching model, see Albert and Chib, 1993). In practice, then, applied economists interested in a thorough data analysis involving a wide variety of nonlinear and structural break models will be interested in Bayesian methods which can be done analytically or at most require numerical integration over a low dimensional subset of the parameter space. This, of course, places restrictions on the types of models analyzed and the priors used.

This leads us to the second possible drawback of the Bayesian approach: the use of proper informative priors. It is well-known in the context of nested hypothesis testing that improper 'noninformative' priors over the parameters of interest lead to Bayes factors which always prefer the restricted model (see Poirier, 1995, section 9.10). Since the Bayes factor is the ratio of the marginal likelihood of the linear model to the marginal likelihood of the alternative model in our examples, the use of the standard uninformative priors will lead to all the posterior probability being in favor of the linear model. By continuity, relatively flat priors over the parameter space will thus tend to favour the restricted (i.e. linear) model. Koop and Potter (1996) argue that this is an attractive feature of the Bayesian analysis in this context since it incorporates a strong reward for the relative parsimony of linear models. Although it is probably unrealistic to expect researchers to universally embrace the use of informative priors, we are confident that those used in this paper will be thought reasonable by a wide variety of readers. Furthermore, optimal classical tests also use a subjective weighting function (see Andrews and Ploberger, 1994).

1.3 Two Examples

We apply our methods to two commonly-analyzed data sets. The first is a post-war quarterly US real GDP growth series. We use CITIBASE series GDP from 1954:1 to 1995:1 in 1987 prices. A number of papers have found evidence of nonlinearity in this data set (see Pesaran and Potter, 1997 for a review) and there is considerable debate about the possibility of a structural break in the time series in the early 1970s (see Perron 1989). The second is a long annual UK industrial production growth series. This latter series runs from 1700-1992 and has been extensively investigated by economic historians (see Greasley and Oxley, 1994, and Mills and Crafts, 1996) who examine whether or not the industrial revolution was a distinct epoch (i.e. whether a structural break occurred in this series).

We work with a similar, relatively noninformative, prior for both data sets. We give examples of how the priors could be altered in our discussion of the models and techniques required to calculate the marginal likelihood.

2 Tools for Analyzing Switching Regime Models

2.1 A Simple Class of Regime Switching Models

Most of the nonlinear models commonly used by macroeconomists are based on autoregressive specifications (i.e., there are no moving average components) which vary across regimes or states. Prominent examples include Markov switching (Hamilton, 1989) and Threshold autoregressive (or TAR, see Potter, 1995) models. In this paper we use the latter class of models since computationally they are much simpler to deal with. Structural break models can also be interpreted as switching regime models where the regimes are defined by an fixed but unknown breakpoint in time. Similarly in our specification outlier models have temporary regimes that occur for one period only.

These considerations give the following simple regime switching specification:

$$Y_t = \begin{cases} \phi_{10} + \phi_{1p}(L)Y_{t-1} + \sigma_1 V_t & \text{if } I_t = 1 \\ \phi_{00} + \phi_{0p}(L)Y_{t-1} + \sigma_0 V_t & \text{if } I_t = 0 \\ \phi_{20} + \phi_{2p}(L)Y_{t-1} + \sigma_2 V_t & \text{if } I_t = 2 \end{cases}$$

where I_t is an indicator variable for the regimes and $\phi_{ip}(L)$ is a polynomial of order p in the lag operator.[5] V_t is assumed to be standard normal and independent over time. The Gaussianity assumption is strong but it will allow us to obtain analytical results.

We consider four ways of defining I_t.

1. The linear Gaussian AR model is obtained if we set $I_t = 0$ for all t.

2. A three regime TAR model is obtained if we set $I_t = 1$ if $Y_{t-d} > r_1$, $I_t = 2$ if $Y_{t-d} < r_2$ and $I_t = 0$ if $r_2 \leq Y_{t-d} \leq r_1$. If $\phi_{20} = \phi_{00}$, $\phi_{0p}(L) = \phi_{2p}(L)$, $\sigma_0 = \sigma_2$ then a two regime model is obtained

3. The structural break model is obtained if we set $I_t = 0$ if $t < \tau_1$ and $I_t = 1$ if $\tau_1 \leq t < \tau_2$ and $I_t = 2$ if $t \geq \tau_2$. Again under suitable parameter restrictions it reduces to a single break model.

4. The outlier model is obtained if $I_t \neq 0$ for only two values of t and $\phi_{1p}(L) = \phi_{2p}(L) = \phi_{0p}(L)$, $\sigma_1 = \sigma_2 = \sigma_0$ but $\phi_{10} \neq \phi_{20} \neq \phi_{00}$.

Of course, there are many other ways of defining I_t and, hence, the tools developed in this paper can be used in a much wider class of models. Note that these models are potentially heteroscedastic since the error variance may differ across regimes. In the empirical section of the paper, we also consider homoscedastic versions of the TAR and structural break models. That is, our homoscedastic models assume $\sigma_1 = \sigma_2 = \sigma_0$. For future reference, let $\phi_i = (\phi_{10}, \phi_{1p}')', (i = 0, 1, 2)$, $\tau = [\tau_1, \tau_2]'$ and $\gamma = [r_1, r_2, d]'$.

[5] For the sake of simplicity we assume the lag length is the same in each regime – an assumption that be easily relaxed.

In constructing a classical test for the presence of a structural break or outlier, the null hypothesis is the AR model and the parameter vector τ is unidentified. In the TAR, the parameter vector γ is unidentified under the AR null. These are the sources of the classical statistical problems discussed previously. Similarly in testing the two regime models versus three regime models or single break versus two breaks or single outliers versus two outliers, r_2 or τ_2 are unidentified under the null hypothesis.

2.2 Analytical Expressions for Marginal Likelihoods

The major advantage of these simple classes of regime switching models is that analytical expressions for the posterior distribution are available conditional on the parameter vectors γ or τ, if we use independent (across i) natural conjugate priors for ϕ_i, σ_i. Hence, it is feasible to carry out a thorough analysis of model choice (i.e. choice of I_t or even choice of heteroscedasticity vs. homoscedasticity), lag length and prior sensitivity within the natural conjugate class. It is worth stressing that, with typical sample sizes and lag lengths, our approach allows for such an analysis to be done in minutes of computer time on a Pentium PC, whereas other approaches (e.g. classical approaches using the optimal tests of Andrews and Ploberger, 1994 and the simulation approach of Hansen, 1996 or Bayesian analysis of Markov switching models as in Albert and Chib, 1993) would take days or weeks of computer time.

Intuitively, the existence of analytical results is due to the fact that, conditional on knowing I_t, the regime switching model breaks down into three AR specifications. If we treat initial conditions as fixed, it is well known that the AR model can be analyzed in a similar manner to the standard linear regression model. Hence, analytical posterior results exist for parameters of each regime, ϕ_i, σ_i, if we work with natural conjugate priors for each regime. To obtain posterior results which are not conditional on I_t, we need to know the marginal posterior for the parameters which define I_t (i.e. τ for the structural break model and γ for the TAR). For example, in the case of the single break model τ_1 the possible values are defined by the integers from 1 to T. Since these parameters are discrete, it is simple to calculate their marginal probability for each possible value. In the case of the TAR, $r_1$and r_2 are continuous parameters but their effect on the likelihood function is the same as if they were discrete. This is because one can set the possible values of r_1, r_2 to the observed data points. For example, suppose r_2 is chosen to be the 25th smallest observation. Then as we vary r_2 up to the value of the 26th smallest observation, the regime split between regime 0 and 2 remains the same. Thus the likelihood function is flat between these values and the posterior for (ϕ_2, σ_2) is the same across all these values for r_2.

Included in the class of priors for which analytical results are available are the improper priors: $p(\phi_i) \propto c_i$, $p(\sigma_i) \propto 1/\sigma_i$, where $c_i > 0$ is an arbitrary constant. As discussed above these priors will lead to all the posterior probability being in favor of the linear model. Thus, we assume an informative prior of the form

$$p(\phi_0, \phi_1, \phi_2, \sigma_0, \sigma_1, \sigma_2|\gamma) = p(\phi_0, \phi_1, \phi_2, \sigma_0, \sigma_1, \sigma_2|\tau) = \prod_{i=1}^{3} p(\phi_i|\sigma_i)p(\sigma_i),$$

with a normal-inverted gamma form for $p(\phi_i|\sigma_i)p(\sigma_i)$ (see, eg., Judge, Griffiths, Hill, Lutkepohl and Lee, 1985, pp. 106-107 for further details about the normal-inverted gamma prior). Thus, conditional on σ_i^2 it is assumed that ϕ_i is multivariate normal with mean vector $\underline{\phi}$ and variance covariance matrix $\sigma_i^2\underline{\mathbf{D}}$. The degrees of freedom of the inverted gamma distribution for the precision is $\underline{\nu}$ and the mean is $\underline{s}^{-2}$. It is assumed that the hyperparameters of the prior do not depend on γ or τ. This is a restriction that could be relaxed with little difficulty. The priors for γ and τ are discussed below.

Let us begin by considering the linear AR model. If we condition on p initial values of Y_t, then this model is identical to the standard Normal regression model with a natural conjugate prior. The analytical expression for the marginal likelihood is well-known (eg. Judge, Griffths, Hill, Lutkepohl and Lee, 1985, page 129). The following sample information is required to calculate the marginal likelihood:

1. The sample size T

2. The ordinary least squares estimates of the parameter vector ϕ:

$$\hat{\phi} = [X'X]^{-1}X'Y,$$

   where $Y = [Y_{p+1}, \cdots, Y_{T+p}]'$ and

$$X = \begin{bmatrix} 1 & Y_p & \cdots & Y_1 \\ 1 & Y_{p+1} & \cdots & Y_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{T+p-1} & \cdots & Y_T \end{bmatrix}.$$

3. The moment matrix $X'X$.

4. The sum of squared errors:

$$e^2 = (Y - X\hat{\phi})'(Y - X\hat{\phi}).$$

Combining the sample information with the prior information, the marginal likelihood is:

$$\ell_{LAR}(Y) = \frac{,\ (\bar{\nu}/2)(\underline{\nu}\underline{s}^2)^{\underline{\nu}/2}}{,\ (\underline{\nu}/2)\pi^{T/2}} \frac{|\bar{\mathbf{D}}|^{1/2}}{|\underline{\mathbf{D}}|^{1/2}} (\bar{\nu}\bar{s}^2)^{-\bar{\nu}/2},$$

where:

1.

$$\bar{\nu} = \underline{\nu} + T$$

2.

$$\bar{\nu}\bar{s}^2 = \underline{\nu}\underline{s}^2 + e^2 + (\bar{\phi} - \hat{\phi})'X'X(\bar{\phi} - \hat{\phi}) + (\bar{\phi} - \underline{\phi})'\underline{\mathbf{D}}^{-1}(\bar{\phi} - \underline{\phi}),$$

   where $\bar{\phi}$ is the posterior mean of ϕ and is given by:

$$\bar{\mathbf{D}}[\underline{\mathbf{D}}^{-1}\underline{\phi} + X'y].$$

3. $\bar{\mathbf{D}}$ combines the prior and sample information on the variance covariance matrix of the lagged variables and is given by:

$$\bar{\mathbf{D}} = [\underline{\mathbf{D}}^{-1} + X'X]^{-1}.$$

Conditional on τ, the homoscedastic structural break and conditional on γ, the homoscedastic TAR models can be written (using appropriate dummy variables) in the form of a linear regression model. Furthermore, conditional on τ the heteroscedastic structural break and conditional on γ the heteroscedastic TAR models divide into three simple linear regression models, each of which has a known marginal likelihood. Hence, conditional on γ or τ the marginal likelihood for these models can easily be calculated. The following sample information is required to calculate the marginal likelihood for the two regime model:

1. The sample size T_i for each of the regimes in the case of the heteroscedastic model.

2. The ordinary least squares estimates of the parameter vector ϕ_i:

$$\hat{\phi}_i = [X_i'X_i]^{-1}X_i'Y_i,$$

where $Y_i = 1(I_t = i)Y$ and $X_i = 1(I_t = i)X$.

3. The moment matrices $X_i'X_i$.

4. The sum of squared errors within each regime:

$$e_i^2 = (Y_i - X_i\hat{\phi})'(Y_i - X_i\hat{\phi}).$$

5. In the case of the heteroscedastic model the 3 pieces of sample information are used to construct separate $\bar{\nu}_i\bar{s}_i^2$, , $(\bar{\nu}_i/2)$ and $|\bar{\mathbf{D}}_i|$ values. In the case of the homoscedastic model the information is combined in constructing the values of $\bar{\nu}\bar{s}^2$ and , $(\bar{\nu}/2)$ and $|\bar{\mathbf{D}}|$.

6. Efficient algorithms can be constructed to calculate this sample information across all the possible splits of the data.

In the case of two regime heteroscedastic model the marginal likelihood conditional on γ or τ is:

$$\ell_{2REGIME}(Y|\gamma) = \prod_{i=0}^{1} \frac{\Gamma(\bar{\nu}_i/2)(\underline{\nu}\underline{s}^2)^{\underline{\nu}/2}}{\Gamma(\underline{\nu}/2)\pi^{T_i/2}} \frac{|\bar{\mathbf{D}}_i|^{1/2}}{|\underline{\mathbf{D}}|^{1/2}} (\bar{\nu}_i \bar{s}_i^2)^{-\bar{\nu}_i/2}.$$

A comparison of the marginal likelihoods for the linear AR and the two regime heteroscedastic model shows that:

1. $\prod_{i=0}^{1}(\bar{\nu}_i \bar{s}_i^2)^{-\bar{\nu}_i/2}$ is likely to be larger than $(\bar{\nu}\bar{s}^2)^{-\bar{\nu}/2}$ even if

   $$\bar{\nu}\bar{s}^2 \approx \bar{\nu}_0 \bar{s}_0^2 + \bar{\nu}_1 \bar{s}_1^2$$

   but this is balanced by the fact that

   $$\Gamma(\bar{\nu}/2) > \prod_{i=0}^{1} \Gamma(\bar{\nu}_i/2),$$

   when $T > 1$ and $T_i > 0, i = 0, 1$.

2. Parsimony in terms of the number of conditional mean parameters used in the model comes from the fact that it is very likely that $|\bar{\mathbf{D}}| > |\bar{\mathbf{D}}_0||\bar{\mathbf{D}}_1|$. For example, in the case of $p = 0$ and assuming that $\underline{\mathbf{D}}$ was 1 we would have $|\bar{\mathbf{D}}| = 1/(T+1)$ and

   $$|\bar{\mathbf{D}}_0||\bar{\mathbf{D}}_1| = \frac{1}{(T_0+1)(T_1+1)},$$

   with $T_0 + T_1 = T$.

Similar considerations apply to the comparison of two regime models with three regime models and three regime models with linear models.

For the comparison across two regime models some insight can be gained into the threshold versus structural break case. Suppose that the sample size of each regime is the same then the ratio of the marginal likelihoods would have two components: the relative size of $\prod_{i=0}^{1}(\bar{\nu}_i \bar{s}_i^2)$ and the relative size of $|\bar{\mathbf{D}}_0||\bar{\mathbf{D}}_1|$ between the two models. If we assume that the latter is approximately equal across models, then the comparison will mainly be determined by the relative fit of the two models. Recall though, that this is still only a conditional comparison and its effect on the overall Bayes factor for the two models will depend on the weight the relevant value of γ and τ receives.

For the structural break model, a discrete uniform prior over all possible sample breaks which imply at least 15% of the data lies in each regime is used for τ. Thus, we do not use any *a priori* information about the location of the break point. The 15% rule is intended to ensure that an adequate amount of data is available in each regime and is consistent with the notion that structural breaks are spaced out over time. The marginal posterior $p(\tau|Y)$ is easy to derive from these conditional marginal likelihoods by normalizing them to a probability measure. The marginal likelihood for the model is found by simply averaging the conditional marginal likelihoods across all the values of τ.

For the TAR we use a continuous uniform prior over r_1, r_2 again with the restriction that at least 15% of the data must be in each regime. This leads to the following prior: r_1 has a uniform distribution from the 30th percentile of the data to the 85th percentile; conditional on r_1, r_2 is uniformly distribution with lower support at the 15th percentile and upper support at the sample value with 15% of the data below r_1. The prior for d is discrete uniform over the integers $1, 2, \ldots, p$. This means that the possible values of d depend the autoregressive lag used. This is an assumption that could also be relaxed.

The calculation of the overall marginal likelihood in the nonlinear case is a little more difficult. Again for each different value of γ we have calculated $\ell(Y|\gamma)$. First, for each fixed pair of r_1, r_2 we can average out the discrete values of d. This gives us $\ell(Y|r_1, r_2)$. Since the likelihood is flat for values of thresholds between the data points this is also true of the marginal likelihood. Thus to integrate the conditional marginal likelihoods against the uniform prior on the thresholds one needs to weight the conditional marginal likelihoods by the size of the interval and divide by the height of the integrating constant of the prior. In the case of the three regime model this achieved by first averaging out over r_2 conditional on r_1 (note that this requires a different integrating constant as r_1 varies). If the threshold effect is very obvious, that is, clear breaks can be observed in the path of the time series, then the conditional marginal likelihoods around the observed thresholds will receive additional weight over other points. This is because for the threshold effect to be visually obvious the marginal distribution of the data will be multimodal with large gaps between the modes.

2.3 Models and Hyperparameters

The results of the previous paragraphs can be extended in the obvious way to cases where more than two regime, structural break or outlier occurs. This paper limits itself to the following 11 classes of models (short form acronyms are given in bold):

1. Linear autoregressive **LAR**.
2. Homoscedastic TAR with one threshold and, hence, two regimes **TAR2-hom**.
3. Heteroscedastic TAR with one threshold and two regimes **TAR2-het.**
4. Homoscedastic TAR with two thresholds and, hence, three regimes **TAR3-hom**.
5. Heteroscedastic TAR with two thresholds **TAR3-het**.
6. Homoscedastic structural break model with one break **Break1-hom**.
7. Heteroscedastic structural break model with one break **Break1-het**.
8. Homoscedastic model with two structural breaks **Break2-hom.**
9. Heteroscedastic model with two structural breaks **Break2-het.**

10. Outlier models with one outlier **Out1.**

11. Outlier models with two outliers **Out2**.

We also examine the models across various lag lengths. For example, in the case of the industrial production data, we let p take on values between 1 and 6. Hence, we compare 66 models in that case.

It remains to discuss the choice of the hyperparameters for the Normal-inverted gamma priors. We begin by eliciting a prior which is relatively noninformative but accords with our subjective prior beliefs. To simplify matters, we assume the prior means and covariances are zero for all the regression parameters in all models (i.e. the ϕ_i's are centered over zero). The prior variances for the regression coefficients are assumed to be the same for all parameters except the intercept(s) for all models (in the case of outlier models the prior variance is the same as the slope coefficients). The prior variance for the intercept is taken to be 10 times as large as that for the slope coefficients. In particular, we assume the marginal prior variance for ϕ_i is $E(\sigma_i^2)cA_{p+1}$ where $A_{p+1}$is a $(p+1)\times(p+1)$ diagonal matrix with $(1,1)'th$ element 10 and all other diagonal elements 1. Hence in the notation above $\underline{\mathbf{D}} = cA_{p+1}$. Degrees of freedom ($\underline{\nu}$) for the inverted gamma priors are 3 for all models, which is very noninformative, but which allows for the first two marginal prior moments to exist for all parameters. The other hyperparameter of the inverted gamma prior is $\underline{s}$. This hyperparameter is defined so that $E(\sigma_i^2) = \frac{\underline{\nu}}{\underline{\nu}-2}\underline{s}^2$.

For real GDP growth we choose $c = \frac{2}{3}$ and $\underline{s}^2 = 1/4$. This implies a very flat prior for σ_i^2, but one that has mean 0.75. Since the data are measured as percentage changes (e.g. 1.0 implies a 1.0 percent change in GDP), this choice of $\underline{s}^2$ is sensible. The prior centers the AR coefficients over zero for all series,[6] but allows for great prior uncertainty. In particular, prior variances of the AR coefficients are 0.5 which implies very large prior

[6]Since we are working with differenced data, this implies a prior centered over a random walk for the level of the series.

standard deviations relative to the size of the stationary region for AR models.

For growth in real industrial production we expect the error variance to be much larger since this data is observed annually and industrial production tends to be more volatile than GDP. In particular, we choose $\underline{s}^2 = 4/3$ which implies we expect the error standard deviation to be around 4. With this change, we set $c = .25$ which implies the prior variance for the AR coefficients is again 0.5.

3 Empirical Results

US real GDP and UK industrial production are plotted in Figures 1a and 1b respectively. It can be seen that U.K. industrial production is much more volatile than U.S. real GDP. For the former, annual increases/decreases of more than ten percent occur almost five percent of the time, while for the latter increases/decreases of more than two percent are rare. This, of course, is mostly due to the fact that industrial production is observed annually, while GDP is observed quarterly. Further, we note that our prior reflects this difference in volatility. A visual inspection of Figure 1a indicates a possible decline in volatility since the early 1980's and Figure 1b indicates a couple periods of high volatility in the early 1700's and the 1920's and 1930's. However, it is hard to ascertain which of the models is supported by the data through visual inspection, and so we turn to our Bayesian methods.

For both of the series we consider, we selected a maximum value of p which we felt was large enough to capture the dynamics of the data. For the real GDP growth series we chose a maximum value of $p = 4$, while for the industrial productions series we chose a maximum value of $p = 6$. Tables 1 and 2 present posterior model probabilities corresponding to each value of p and each class of models using the prior described in the previous section for US GDP and UK Industrial Production respectively.

Note first that the strong reward for parsimony reflected in Bayes factors with relatively flat priors implies that short lag lengths tend to be preferred for both series. Despite this,

Table 1: Posterior Model Probabilities for real US GDP Data

Model/AR Order	1	2	3	4
AR	0.0007	0.0001	0.0000	0.0000
TAR2-hom	0.0001	0.0000	0.0000	0.0000
TAR2-het	0.0001	0.0062	0.0001	0.0000
Break1-hom	0.0001	0.0000	0.0000	0.0000
Break1-het	0.8011	0.1306	0.0069	0.0003
TAR3-hom	0.0000	0.0000	0.0000	0.0000
TAR3-het	0.0000	0.0007	0.0000	0.0000
Break2-hom	0.0000	0.0000	0.0000	0.0000
Break2-het	0.0481	0.0026	0.0000	0.0000
Out1	0.0008	0.0002	0.0000	0.0000
Out2	0.0008	0.0003	0.0000	0.0000

there is little evidence for the linear model (which, for a given lag length, is the most parsimonious model). This reinforces the common finding that departures from linearity are present in macroeconomic time series. However, just how this departure manifests itself seems to differ over series.

The real GDP growth series seems to be characterized by structural breaks in the variance. The heteroscedastic model with one break gets most of the posterior model probability, although substantial probability is allocated to the two-break model. Another point worth stressing is that the heteroscedastic TAR's tend to be favoured over the linear model (see, especially, TAR2-het with $p = 2$ which is roughly ten times as likely as the linear AR with any choice of p). Hence, if a researcher was just looking for nonlinearity he/she would likely conclude that it is present even though it is likely that this merely reflects a structural break in the volatility of GDP growth. Estimates of the breakpoint indicate that volatility of the innovation to US GDP changed markedly in the early- to mid-1980's. Note also that there is little evidence for any nonlinearities in the conditional mean of the series (i.e., the homoscedastic models do not beat the linear AR model by

Table 2: Posterior Model Probabilities for UK Industrial Production Data

Model/AR Order	1	2	3	4	5	6
AR	0.0055	0.0039	0.0005	0.0000	0.0000	0.0000
TAR2-hom	0.1215	0.0176	0.0085	0.0003	0.0000	0.0000
TAR2-het	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Break1-hom	0.0090	0.0070	0.0005	0.0001	0.0000	0.0000
Break1-het	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TAR3-hom	0.1363	0.2797	0.0486	0.1922	0.0036	0.0001
TAR3-het	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Break2-hom	0.0598	0.0839	0.0016	0.0000	0.0000	0.0000
Break2-het	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Out1	0.0054	0.0038	0.0005	0.0000	0.0000	0.0000
Out2	0.0057	0.0039	0.0005	0.0000	0.0000	0.0000

much when limiting the comparison to these two groups). An additional interesting point is that, if we restrict attention to homoscedastic models, the outlier models are the most preferred (although they only slightly beat the linear AR model).

Results for the growth in UK industrial production series is markedly different. For this series there is absolutely no evidence for any type of heteroscedasticity, but there is strong evidence for departures from linearity in the conditional mean of the series. The posterior probability is spread out over a lot of different models, but overall the homoscedastic TAR models receive most support. If we sum across p to get the total probability for each class of models, then TAR3-hom receives 67 per cent and TAR2-hom receives 15 per cent of the posterior model probability. So there is surprisingly strong evidence that the departures from linearity observed in this series are not due to structural breaks (despite the length of the time series which includes the industrial revolution), but rather by some endogenous process where changes in linear structure are predictable based on past information. However, we do not want push this story too hard, since there is some evidence for structural change (i.e. the total probability allocated to the Break1-hom and Break2-hom classes of models is around 14 per cent). Furthermore, our results do not completely rule out the possibility that the linear model adequately characterizes the data (i.e. this class gets around 2 per cent of the posterior model probability).

These empirical results are meant to illustrate the practical usefulness of our techniques and show how great care needs to be taken when modelling departures from linearity. Even with these two series we can see that generalizations like: 'structural breaks are present in macroeconomic time series' or 'macroeconomic time series are nonlinear' are misleading. In a more extensive empirical exercise, a researcher would want to carry out a prior sensitivity analysis[7] and present other posterior features of interest. We have not done so for reasons of brevity. Furthermore, in this paper we have emphasized model choice. However, in practice, with Bayesian techniques the choice of one preferred model is not necessary. Rather, features of interest (eg. impulse responses or forecasts) can be presented which average over all possible models where the weights are given by the posterior model probabilities.

4 Conclusions

This paper argues for the use of Bayesian 'tests' as a complement to the myriad of classical tests now extant for testing for nonlinearities or structural breaks. Further, given that macroeconomists usually want to test a wide variety of specifications, it is important to develop tests that can be applied jointly across the specifications. The main methodological point of this paper has been to show how Bayesian methods for handling autoregressive models with unknown changes in regime can be used to construct such a joint test across specifications. Because these methods can be performed analytically, they are both computationally inexpensive and easy to use in practice.

[7] We have informally experimented with changing the prior and, in practice, results seem to be robust to reasonable changes in our present prior.

In the empirical section of the paper, our methods are implemented for two representative macroeconomic time series. For real GDP growth, our findings reinforce those of Stock and Watson (1996): structural breaks do indeed appear to occur. If this finding were widespread in macroeconomics, then a good portion of empirical macroeconomics is called into question. If changes in structure are widespread and unpredictable, then it is difficult to imagine an appropriate forecasting strategy for use with macro data and it may be difficult to develop economic theory to explain this. However, for our industrial production series, departures from linearity seem to be characterized by endogenous changes in structure of the sort captured by the TAR. If this behaviour is widespread in macroeconomics, then a very productive research strategy exists involving development of more sophisticated nonlinear models and theories to explain the nonlinearities.

One obvious extension of our results would be to conduct a formal sensitivity analysis for the hyperparameters $c, \underline{s}^2$. Another important extension would involve examining the sensitivity of the results to the assumption of Gaussian innovations to the time series. Indeed, allowing for Student-t errors might make the outlier models considered in this paper redundant (see, eg., Hoek, Lucas and van Dijk, 1995). However, this would involve some simulation. Thus, to reduce the computational burden it would make sense to focus on a subset of models suggested by an initial analysis based on the computationally simpler techniques described here.

References

[1] Albert, J. and Chib, S. (1993). "Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts," *Journal of Business and Economic Statistics*, 11, 1-15.

[2] Andrews, D.W.K. (1993). "Tests for parameter instability and structural change with an unknown change point," *Econometrica,* 61, 821-856.

[3] Andrews, D. and Ploberger, W. (1994). "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica,* 62, 1383-1414.

[4] Barry, D. and Hartigan, J. (1993). "A Bayesian analysis for change point problems," *Journal of the American Statistical Association,* 88, 309-319.

[5] Beaudry, P., and Koop, G., (1993): "Do Recessions Permanantly Affect Output?" *Journal of Monetary Economics*, 31, 149-163.

[6] Carlin, B., Gelfand, A. and Smith, A.F.M. (1992). "Hierarchical Bayesian analysis of changepoint problems," *Applied Statistics,* 41, 389-405.

[7] Greasley, D. and Oxley, L. (1994). "Rehabilitation sustained: The Industrial Revolution as a macroeconomic epoch," *Economic History Review, 2nd Series*, 47, 760-768.

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

[9] Hansen, B. (1996). "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413-430.

[10] Hoek, H., Lucas, A. and van Dijk, H. (1995). "Classical and Bayesian aspects of robust unit root inference," *Journal of Econometrics*, 69, 27-59.

[11] Judge, G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T.C. (1985). *The Theory and Practice of Econometrics,* second edition, Wiley, New York.

[12] Koop, G. and Potter, S. (1996). "Bayes factors and nonlinearity: Evidence from economic time series," manuscript.

[13] Mills, T. and Crafts, N. (1996). "Trend growth in British industrial output, 1700-1913: A reappraisal," *Explorations in Economic History*, 33, 277-295.

[14] Pesaran, M.H., and Potter, S. (1997). "A Floor and Ceiling Model of US Output," *Journal of Economic Dynamics and Control* 21, 661-695.

[15] Poirier, D. (1995). *Intermediate Statistics and Econometrics,* The MIT Press, Cambridge.

[16] Potter, S. (1995). “A nonlinear approach to US GNP,” *Journal of Applied Econometrics*, 10, 109-126.

[17] Stephens, D.A. (1994). “Bayesian retrospective multiple-changepoint identification,” *Applied Statistics,* 43, 159-178.

[18] Stock, J. and Watson, M. (1996). “Evidence on structural instability in macroeconomic time series relations,” *Journal of Business and Economic Statistics,* 14, 11-30.