

# ARE APPARENT FINDINGS OF NONLINEARITY DUE TO STRUCTURAL INSTABILITY IN ECONOMIC TIME SERIES?<sup>⌘</sup>

GARY KOOP

Department of Economics, University of Edinburgh,  
Edinburgh, U.K. EH8 9JY  
E-mail: G.Koop@ed.ac.uk

SIMON M. POTTER

Domestic Research Function, Federal Reserve Bank of New York,  
33 Liberty St., New York, NY 10045-0001, U.S.A.  
E-mail: simon.potter@ny.frb.org

December 1998

**ABSTRACT:** Many modelling issues and policy debates in macroeconomics depend on whether macroeconomic time series are best characterized as linear or nonlinear. If departures from linearity exist, it is important to know whether these are endogenously generated (as in, e.g., a threshold autoregressive model) or whether they merely re°ect changing structure over time. In this paper, we discuss a model comparison methodology which addresses these issues. We advocate a Bayesian approach and show how such an approach can be implemented in practice. An empirical exercise involving several macroeconomic time series shows that apparent findings of threshold type nonlinearities could be due to structural instability.

Keywords: Bayes Factor, Markov chain Monte Carlo,  
threshold autoregressive model, time varying parameter model  
JEL: C22, C11, E30

---

<sup>⌘</sup>The views expressed in the paper are those of the authors and do not necessarily re°ect the views of the Federal Reserve Bank of New York or the Federal Reserve System.

# 1 Introduction

Much of the empirical macroeconomic literature has been dominated by linear reduced form models (e.g. AR or VAR models). Recently, there has been a great deal of interest in freeing up the symmetry restrictions implicit in linear time series models. For instance, one may wish to examine whether the persistence of shocks is different in recessions and expansions. Much of this work has considered threshold type nonlinear specifications and found extensive, but not overwhelming, evidence for nonlinearity in many key macroeconomic time series (see, among many others, Terasvirta and Anderson, 1992, Potter, 1995, Beaudry and Koop, 1993, and Pesaran and Potter, 1997). However, there is an even larger literature that searches for evidence of structural instability. For example, Stock and Watson (1996) find evidence for structural instability in a wide range of series. Thus, in addition to the question of the importance of these departures from linearity there is the question of whether the apparent evidence for nonlinearities is just evidence that there are structural instabilities in time series relationships that linear fixed parameter models cannot account for.

The importance to macroeconomics of answering these questions cannot be overemphasized. If linear models provide an adequate representation of most economic time series then standard methodologies (e.g. VAR methods) are suitable, and the stylized facts they produce (e.g. the widespread finding of unit roots with the implication that shocks have permanent effects) provide a good basis for policymaking and theorizing. If important nonlinearities exist, however, then these should be incorporated in macro models and the stylized facts generated by linear models are called into question. On the other hand, if widespread structural instability exists in most macroeconomic aggregates then empirical analysis of macroeconomics using fixed parameter time series methods is a dubious exercise at best.

In this paper, we describe a Bayesian method for testing and estimation in light of the issues raised above. We consider the common case where the researcher is considering a certain nonlinear model (e.g. a threshold autoregressive, TAR, or a Markov switching model) that is felt to accommodate potentially important departures from linearity. In order to convince a wide audience of the usefulness of the nonlinear model, the researcher must, of course, provide convincing evidence of departures from linearity. In addition,

the author must convince the reader that apparent nonlinearities do not merely reflect structural change. Hence, we argue that the nonlinear time series econometrician must compare the nonlinear<sup>2</sup> model both to a linear model and to a model containing structural instability.

There are many ways of empirically implementing the general strategy outlined above. In this paper, we base inference on variants and extensions of autoregressive models. We choose to focus on autoregressive (AR hereafter) models because of their flexibility, computational simplicity and common usage in the literature. We reason as follows: an AR model assumes that there is an invariant linear structure which holds at all times; nonlinear extensions of AR models (e.g. TAR models) assume a linear structure which changes in a way predicted by the past history of the time series. This suggests, when choosing a model which reflects structural instability, that we use a time-varying parameter (TVP) AR model where the coefficients evolve over time according to a random walk. Such a model would have a linear structure at any point in time, but this structure will change in a way that is not predictable from the past history of the time series. By basing our intuition on the degree of predictability of the linear structure, we can frame a nonlinear model as lying between two extremes: one where perfect predictability exists (i.e. the linear model where no change in structure occurs) and one where changes in structure are totally unpredictable.

In previous work (Koop and Potter (1999)), we recommend the use of Bayes factors for comparing linear to nonlinear time series models: Bayes factors surmount Davies' problem (i.e. nuisance parameters unidentified under the null) which plagues classical testing; have an Occam's razor property which is important in light of Lindley's paradox (see Poirier (1995), p. 545); allow for easy comparison of non-nested models and allow for pooling inferences from many models in the usual case where more than one model receives sizeable support from the data. In the present work, we also adopt a Bayesian approach. For most of our models, analytical tools can be used to calculate the marginal likelihoods that are necessary to calculate Bayes factors or posterior model probabilities. For the time-varying parameter model, we show how Markov chain Monte Carlo (MCMC) methods can be used to calculate the marginal likelihood.

We use our methods to investigate the presence of nonlinearities in several artificial and real data series.

---

<sup>2</sup>Formally speaking, a structurally unstable model is "nonlinear". However, we will follow common practice and use the term "nonlinear" time series model to refer to models such as the TAR where the change in linear structure over time is at least partially predictable.

Our empirical findings indicate little support for threshold-type nonlinearities alone. Postwar quarterly US GDP growth does not provide strong evidence for departures from linearity. Postwar quarterly US inflation appears to exhibit substantial structural instability as does a long annual UK industrial production series. All of these series could be incorrectly misclassified as having threshold nonlinearities using traditional approaches. We argue that our findings indicate both the importance of Bayesian methods (which lessen the chances of data mining) and of comparing nonlinear models to both linear and structurally unstable models (instead of just the former).

## 2 Models

The models considered in this paper are extensions of the standard AR(p) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \epsilon_t \quad (1)$$

where  $\epsilon_t \gg$  I.I.D.  $N(0,1)$ . We label this  $M_L$  for linear model. It is parameterized by  $\alpha = (\alpha_0; \dots; \alpha_p)'$  and  $\epsilon_t^2$ .

A simple way of extending this model is to allow for AR dynamics to differ across various regimes, where the regimes are defined endogenously using past information in the system. There are many ways in which one can do this, the simplest being the threshold autoregressive (TAR) model (see Potter, 1995) which we will label  $M_{TAR}$ . In this paper, we will focus on a two regime model:

$$y_t = \begin{cases} \frac{1}{2} (\alpha_{01} + \alpha_{11} y_{t-1} + \dots + \alpha_{p1} y_{t-p_1} + \epsilon_t) & \text{if } X_{t;d} > r; \\ \frac{1}{2} (\alpha_{02} + \alpha_{12} y_{t-1} + \dots + \alpha_{p2} y_{t-p_2} + \epsilon_t) & \text{if } X_{t;d} \leq r; \end{cases} \quad (2)$$

Notice that the TAR allows for different dynamics depending on whether a variable  $X_{t;d}$  is above or below a threshold  $r$ .  $X_{t;d}$  is some function of data observed between periods  $t-j-1$  and  $t-j-d$  inclusive. We choose  $X_{t;d}$  to be lags of the dependent variable, i.e.  $X_{t;d} = y_{t-d}$ .  $r$  and  $d$  are treated as unknown parameters in our analysis. To keep the number of models and parameters manageable, we set  $p_1 = p_2 = p$  and  $d = p$ .<sup>3</sup>

The TAR can be reparameterized using dummy variables as:

---

<sup>3</sup>The TAR could easily be extended to allow for heteroskedasticity. We do not do so in order to focus on nonlinearities in the conditional mean of the series.

$$y_t = \phi_0 D_{1t} + \phi_1 y_{t-1} D_{1t} + \dots + \phi_p y_{t-p} D_{1t} + \theta_0 D_{2t} + \theta_1 y_{t-1} D_{2t} + \dots + \theta_p y_{t-p} D_{2t} + \epsilon_t;$$

where  $D_{1t} = 1$  if  $X_{t,d} \leq r$ , and  $= 0$  otherwise and  $D_{2t} = 1$  if  $X_{t,d} > r$ , and  $= 0$  otherwise. Hence, the parameters of the model are the same as the AR model except that  $r; d$  and  $\theta = (\theta_0; \dots; \theta_p)'$  are added.

The third class of models we consider also allows for the dynamics to differ over time, but does not allow for changes in structure to be predictable using past observables. This is a time varying parameter model ( $M_{TVP}$ ) which allows the coefficients of a linear AR representation to evolve according to a random walk:

$$y_t = \phi_{0t} + \phi_{1t} y_{t-1} + \dots + \phi_{pt} y_{t-p} + \epsilon_t;$$

where

$$\phi_{it} = \phi_{i;t-1} + \lambda_i v_{it}$$

and  $v_{it}$  is assumed to be independent of  $y_t$  and is distributed I.I.D.  $N(0,1)$  for  $i = 0; \dots; p$ . We define  $\lambda = (\lambda_0; \dots; \lambda_p)'$ : Note that evidence in favor of significant variations over time in the parameters is also evidence of structural instability. If  $\lambda = 0$ ; this model reduces to the linear AR model. Writing  $\phi_{it} = \phi_{i0} + \lambda_i \sum_{j=1}^t v_{ij}$ , it can be seen that the  $\phi_{i0}$ 's ( $i = 1; \dots; p$ ) are analogous to the AR coefficients in  $M_L$ . Hence,  $M_{TVP}$  can be parameterized in terms of  $\phi = (\phi_{00}; \dots; \phi_{p0})'$ ,  $\lambda^2$  and  $\lambda$ :

Within these three classes of models ( $M_L; M_{TAR}; M_{TVP}$ ); we allow for various submodels depending on AR lag length. In our empirical section, we allow the lag length to range from 1 to 4<sup>4</sup>

### 3 Priors and Bayes Factor Calculation

To calculate Bayes factors in the context of comparison of nested models, proper priors must be specified (i.e. the prior must integrate to one). The use of flat (improper) priors typically results in degenerate Bayes factors that always favor the restricted over the unrestricted model regardless of the data. The

<sup>4</sup>Preliminary data analysis indicated that no values of  $p$  greater than 4 received appreciable posterior support for any of our series.

use of informative priors is often criticized by non-Bayesian econometricians. In addition to the usual Bayesian methodological arguments in favor of informative priors, we note that:

i) In the VAR forecasting work of Litterman and Sims, informative priors are found to greatly improve forecasting performance. This occurs despite the fact that the priors are chosen mainly for computational simplicity. Prior hyperparameters are chosen in a fairly crude way using only rough reduced form information (e.g. macroeconomic series are probably roughly I(1), AR coefficients are probably not too large, distant lags of the series are probably less important than recent lags, etc.). That is, prior information can be useful in an empirical modelling exercise even if the information is very inexact.

ii) Optimal solutions to Davies' problem require the elicitation of something that is essentially identical to a prior (see Koop and Potter (1999)), so classical approaches are also dependent on informative priors.

iii) Bayes factors are usually robust to "reasonable" changes in the prior. In practice, we carry out a sensitivity analysis.

iv) In this paper we are interested in comparing different classes of models. These classes have many common parameters (the AR coefficients and error variance). We use the same prior for these parameters in each class of models. It is only for the few parameters that are not common to all models (i.e.  $\alpha$  and  $\beta$ ) that prior information is potentially important. Loosely speaking, this means that it is less likely that the prior is driving the Bayes factors obtained (i.e. since the prior is so similar, it must be likelihood differences driving the Bayes factor results).

Throughout this paper, we condition on  $p$  initial observations and, hence, the AR model has the same likelihood function as the Normal linear regression model.<sup>5</sup> Accordingly, we make use of the Normal-inverted Gamma natural conjugate prior for  $\alpha$  and  $\beta$ . Writing  $p(\alpha; \beta) = p(\alpha|\beta)p(\beta)$ , we take  $p(\alpha|\beta)$  to be  $N(0; \beta^2 A)$  and  $p(\beta)$  to be inverted Gamma with hyperparameters  $\nu_\beta$  and  $\rho_\beta$  (we adopt the notation for the inverted-Gamma given in Judge, Griffiths, Hill, Lutkepohl and Lee (1985), pages 106-107). Prior degrees of freedom are given by  $\nu_\beta$  and the prior mode of  $\beta$  is  $\frac{\rho_\beta}{\nu_\beta + 1}$ . With this prior, the marginal likelihood has an analytical form (see, e.g., Judge, Griffiths, Hill, Lutkepohl and Lee (1985), page 129). This success

<sup>5</sup>From this point forth, we will write the data as running from period 1 to  $T$ , implicitly labelling the first  $p$  observations as 1 through 0.

to describe the likelihood, prior and computational techniques for calculating the marginal likelihood for  $M_L$ :

For  $M_{TAR}$ , we write the prior as:  $p(\theta; \phi; r; d) = p(\theta; \phi) p(\phi) p(r|d) p(d)$ . We use a prior for  $\theta; \phi$  and  $\phi$  which is analogous to that used for  $M_L$ . In particular, we use a Normal-inverted Gamma prior for these parameters where the inverted-Gamma part is identical to that given above. Furthermore, the prior for the regression coefficients, conditional on  $\phi^2$ , is given by:

$$\mu_{\theta} \sim N(0; \phi^2) \quad \Sigma_{\theta} \sim \text{IG}(A, 0) \quad \Sigma_{\phi} \sim \text{IG}(A, 0)$$

Flat priors are used for the parameters  $r$  and  $d$ . In particular,  $r$  is assumed to be a priori flat over all possible values that imply that at least 15% of the observations lie in each regime and  $d$  is flat over the integers  $1; \dots; p$ . Note that the prior is centered over the linear specification. The marginal likelihood for this model can easily be calculated by noting that, conditional on  $r$  and  $d$ ,  $M_{TAR}$  is a linear regression model and standard analytical results can be used. An unconditional marginal likelihood can be calculated by evaluating the conditional marginal likelihood at every possible  $r, d$  combination and then averaging using the priors for these latter two parameters. Koop and Potter (1998a,b) provide further details regarding TAR models.

$M_{TVP}$  contains the same parameters as  $M_L$  plus  $\psi$ . For the common parameters, we use the same prior.  $\psi$  is assumed to be independent of the common parameters. We assume  $p(\psi) = \prod_{i=1}^{p+1} p(\psi_i)$  and  $p(\psi_i)$  is inverted Gamma with hyperparameters  $\alpha_{\psi_i}$  and  $\beta_{\psi_i}$ :

The marginal likelihood for  $M_{TVP}$  is much more difficult to calculate, since analytical results are not available. However, as described below, a Gibbs sampler can be set up to take pseudo-random draws from the posterior for this model. Given the output from this posterior simulator, the method of Chib (1995) can be applied directly to calculate the Bayes factor.

### 3.1 Values for Prior Hyperparameters

In this paper we use either artificially simulated data or real data. In order to simplify prior elicitation for these various series, the artificial data is simulated with  $\phi = 1$  and the real data is in growth rates and

standardized to have mean zero and unit standard deviation. With all these data sets, it is reasonable to use the same prior. The prior hyperparameters are selected to be intuitively reasonable, but fairly noninformative. An appendix contains results from a sensitivity analysis where the prior hyperparameters are varied.

Given that we are using demeaned, growth rate data, it is sensible to choose prior means of zero for all regression coefficients. The prior variance of these coefficients, conditional on  $\phi^2$ , is  $\phi^2 A$ . We set  $A = 0.25 \mathbf{I}$ . Given that  $\phi$  is likely around 1, the prior standard deviation of the regression coefficients will be roughly .5 indicating that we think it unlikely that they are larger than 1 in absolute value. In other words, we are fairly confident (but not dogmatically so) that the AR coefficients lie in the stationary region.

For the error variance, we set  $\underline{\phi} = 1$  and  $\sigma_{\epsilon}^2 = 3$ . Since the degrees of freedom parameter can, loosely speaking, be thought of as a prior sample size, this prior is fairly noninformative (i.e. it has roughly the same information content as three data points) but it is centered near 1.

It is worthwhile to justify the prior for the  $\rho_i$ s in more detail and to distinguish between the case where the data is in levels as opposed to growth rates. We begin by considering the case where  $p=0$  (i.e. the model contains only an intercept which follows a random walk). It is well-known that this model implies that  $y_t$  contains a unit root and that  $\rho_0$  is, loosely speaking, a measure of how large the unit root component is. This forms the basis for several unit root tests (see, e.g., Kwiatkowski, Phillips, Schmidt and Shin, 1992). This suggests that, if  $y_t$  is a growth rate,  $\rho_0$  is likely quite small, since the I(2) component of the original level series is likely quite small. On the other hand, if  $y_t$  is a level series, it is possible that  $\rho_0$  is much larger.

If  $p>0$ , then the evolving parameters include AR coefficients, which are unlikely to change by much in any period. Consider the case where  $p=1$ . If  $\rho_{1t}$  often changes by more than, say, 0.1 in any period it will, with high probability, wander into the explosive region of the parameter space. Allowing for changes of this magnitude is not reasonable for macroeconomic time series unless massive technical change is repeatedly occurring. Such considerations hold regardless of whether  $y_t$  is a growth rate or a levels series. In other

words,  $\alpha_i$  should be very small for  $i=1,\dots,p$ .

These properties are illustrated in Figures 1 and 2, which plot six different artificial series with  $T=200$  and  $p=1$ .<sup>6</sup> All have  $\alpha_{00} = 0$ ,  $\alpha_{10} = 0.5$  and  $\beta = 1$ ; but vary in their choices of  $\alpha_0$  and  $\alpha_1$ . All series are normalized to have zero mean and unit variance. Figure 1 chooses small values for  $\alpha_0$  which are typical of growth rates, while Figure 2 sets  $\alpha_0 = 1$ .  $\alpha_1$  is then varied in each figure. These two figures illustrate three points: i) Setting  $\alpha_1 = 1$  in either figure allows the AR(1) coefficient to wander into explosive regions. This value seems to be too high for macroeconomic time series; ii) If we look at the cases where  $\alpha_1 = 0$ , then we can see how  $\alpha_0$  affects the trend properties of the series; and iii) Moderate values for  $\alpha_1$  such as 0.05 do generate series which look like they could plausibly model real macroeconomic data.

Given these considerations, and the fact that we are using growth rates, we set  $\alpha_i = 1$  and  $\alpha_i = 2$  for  $i = 0, \dots, p$ : This is a very noninformative prior. In particular, it has a median of .12. In other words, this prior allocates roughly half its weight to the region we feel is plausible,  $[0, .1]$ , but has an extremely long tail so that some prior weight is allocated even to implausibly large values of  $\alpha_i$ .

## 4 MCMC Algorithm

The MCMC algorithm for  $M_{TVP}$  can be obtained by modifying the existing literature on Bayesian literature on state space models. That is, conditional on knowing the states (i.e.  $\alpha_i^s = (\alpha_{i1}, \dots, \alpha_{iT})^0$  for  $i = 0, \dots, p$ ) the distribution of the remaining parameters is simple to derive { it becomes a standard linear regression model with natural conjugate prior. Furthermore, conditional on knowing the parameters of the model, random draws of the states can be taken using established methods. This informal logic suggest a Gibbs sampler with data augmentation can be used.

Formally, we can take random draws from  $p(\alpha; \beta | \text{Data}; \alpha, \beta)$ , where  $\alpha^s = (\alpha_0^s; \dots; \alpha_p^s)$ ; using the techniques of de Jong and Shephard (1995). Techniques for drawing from  $p(\alpha; \beta | \text{Data}; \alpha^s)$  are similarly easy since this posterior conditional has the familiar Normal-inverted Gamma form. In particular, remember that for  $M_{TVP}$ ;  $\alpha$  is defined as the initial states in the state equation and that  $\alpha_{it} = \alpha_{i0} + \alpha_i \beta \sum_{j=1}^t V_{it}$ .

<sup>6</sup>All series are created with the same initial seed in the random number generator.

Hence, the model can be written as:

$$y_t^a = \alpha_{00} + \alpha_{10}y_{t-1} + \dots + \alpha_{p0}y_{t-p} + \eta_t^a;$$

where, conditional on knowing the states,  $y_t^a$  is known and is given by:

$$y_t^a = y_{t-1} \alpha_{10} + \sum_{j=1}^p \alpha_{j0} V_{0j} + \sum_{j=1}^p \alpha_{j1} V_{1j} + \dots + \sum_{j=1}^p \alpha_{jp} V_{pj} + \eta_t^a;$$

We choose this non-standard way of setting up the model since it becomes clear that, conditional on knowing the states, the model becomes a linear regression model and, hence, the Normal-inverted Gamma prior for  $(\alpha; \eta)$  is (conditionally) natural conjugate.

It remains to derive the form of  $p(\beta | \text{Data}; \alpha; \eta; \eta)$ . Using the inverted Gamma prior for this parameter discussed in the previous section, it follows that the posterior conditional is also inverted Gamma (see, for instance, Min (1992)).

Details on the Gibbs sampler and the marginal likelihood calculation for the TVP model are given in Appendix A.

## 5 Application to Artificial Data

Before working with real macroeconomic data, it is instructive to consider the performance of our Bayesian methodology in artificial data. Accordingly, in this section we simulate one data set of size 200 from each of 20 different parameter configurations.<sup>7</sup> In order to focus on the comparison between linear, TAR and TVP models we always set  $p=1$  and  $\eta=1$ . The other parameters are set in order to shed light on the following questions:

1. Is it possible for TVP data generating processes to be misclassified as TAR models?
2. Is it possible for TAR data generating processes to be misclassified as TVP models?

Accordingly, we carry out two experiments. In Experiment 1, we generate artificial data sets from  $M_{TVP}$  with  $\alpha_{00} = 0$  and  $\alpha_{01} = .5$  and  $\eta_0 = \eta_1$ . We then vary this common value for  $\eta$  across data sets

<sup>7</sup>The same seed for the random number generator was used for every artificial data set.

over the region  $[0,0.08]$ . In our previous discussion on prior hyperparameter selection we have motivated this as an interesting region of the parameter space. The case  $\rho_0 = \rho_1 = 0$  corresponds to an AR(1) model with coefficient  $\rho_1$ . The case  $\rho_0 = \rho_1 = 0.08$  has a large degree of coefficient variation over time.

In Experiment 2, we generate artificial data sets from  $M_{TAR}$  with  $\mu_0 = \mu_1 = \sigma_0 = 0$ ,  $r=0$  and  $d=1$ . The coefficient  $\rho_1$  is allowed to vary across data sets over  $[0,0.9]$ . If  $\rho_1 = 0$  then the model is linear. However, as  $\rho_1$  increases the dynamics in the two regimes become more and more different.

Figures 3 and 4 present the results of our two experiments in graphical form, using the prior described above. In order to visually highlight the information in these figures we plot the logs of Bayes factors comparing the TAR and TVP models to the linear model. Note that the log of the Bayes factor being zero implies that the models under consideration receive equal support from the data, values greater than zero support the linear model and values less than zero support either the TAR or TVP model.

Figures 3 and 4 indicate that Bayes factors do a reasonable job of supporting the correct model, but have a strong reward for parsimony built in. In Figure 3, if  $\rho_1$  is less than roughly 0.025 (i.e. there is a small degree of structural instability), the linear model is supported over the TVP model. In Figure 4, the difference in AR(1) coefficients across regimes has to be greater than 0.2 for the TAR model to be preferred to the linear model. In other words there has to be a fairly sizeable degree of nonlinearity or structural instability before the Bayes factor will support these more parameter rich models.

With regards to the two questions posed at the beginning of this section, our results indicate that the answers are Yes and No, respectively. In particular, Figure 3 indicates that if there is a large degree of structural instability, then it is possible for the Bayes factor comparing the linear to the TAR model to indicate substantial support for the latter. If we had omitted to consider  $M_{TVP}$ , we could (incorrectly) have concluded our artificial data sets with  $\rho_1 = 0.07$  or higher were generated from TAR models. In contrast, Figure 4 indicates that there is no way that a threshold model could be mistaken for a structurally unstable model. Even for large asymmetries between dynamics in the two regimes, our methodology always indicates that the linear model beats the TVP model.

These results are, of course, suggestive rather than conclusive. Remember, we only generated one data

set for each of a limited set of parameter configurations. However, they do indicate that including both a linear and a time-varying parameter model to compare with a candidate nonlinear model is potentially of great importance.

## 6 Application to Real Data

The data used in this section is:

1. DGDP: Quarterly real US GDP growth from 1954Q2 through 1995Q1.
2. Inflation: Quarterly change in the US CPI from 1947Q2 through 1998Q3.
3. DIP: Annual change in UK industrial production from 1701-1992.

These three selections were made to reflect the different sorts of series macroeconomists often work with. The DGDP series has been examined in a number of papers, which have found some (weak) evidence of nonlinearity in this data set (see Pesaran and Potter, 1997 or Kapetanios, 1998 for a review). The postwar behavior of inflation indicates that structural instability could be a problem for any time series model of this series. DIP is a long annual UK industrial production growth series. This latter series runs from 1701-1992 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 that reflected a large degree of technical progress.

The data (normalized to have mean zero and variance 1) is plotted in Figure 5. Tables 1 through 3 present posterior model probabilities<sup>8</sup> for each of the 3 models for each the four lag lengths. The three series seem to exhibit quite different behavior, so we discuss each in turn.

Table 1 indicates that there is almost no evidence in favor of structural instability for DGDP. That is, the linear and TAR models receive virtually all of the posterior model probability, as do short lag lengths. Overall, there does not seem to be enough data information to decide between  $M_L$  or  $M_{TAR}$  | both receive appreciable support from the data (roughly 70% and 30%, respectively). However, there definitely does not seem to be the sort of overwhelming evidence which would lead one to abandon the well-understood

---

<sup>8</sup>We attach equal prior weight to each model, so the posterior model probabilities are merely renormalized marginal likelihoods.

Table 1: Posterior Model Probabilities for DGDP

	$M_L$	$M_{TAR}$	$M_{TVP}$	Total
p=1	0.507	0.230	0.005	0.741
p=2	0.151	0.060	0.004	0.215
p=3	0.032	0.005	$1.4 \times 10^{-4}$	0.038
p=4	0.005	0.001	$8.3 \times 10^{-6}$	0.006
Total	0.696	0.296	0.009	1

Table 2: Posterior Model Probabilities for In°ation

	$M_L$	$M_{TAR}$	$M_{TVP}$	Total
p=1	$4.6 \times 10^{i-14}$	$2.2 \times 10^{i-14}$	$1.4 \times 10^{i-10}$	$1.4 \times 10^{i-10}$
p=2	$2.2 \times 10^{i-13}$	$1.5 \times 10^{i-13}$	$1.9 \times 10^{i-9}$	$1.9 \times 10^{i-9}$
p=3	$8.4 \times 10^{i-12}$	$3.9 \times 10^{i-12}$	$7.3 \times 10^{i-8}$	$7.3 \times 10^{i-8}$
p=4	$2.0 \times 10^{i-10}$	$2.1 \times 10^{i-9}$	1.000	1.000
Total	$2.0 \times 10^{i-10}$	$2.1 \times 10^{i-9}$	1.000	1

linear class of models.

Table 2, on the other hand, finds overwhelming evidence of structural instability in inflation.  $M_{TVP}$  receives virtually all of the posterior model probability.<sup>9</sup> A second important finding is that, if we had ignored  $M_{TVP}$  and just compared  $M_L$  and  $M_{TAR}$ , we would have found reasonably strong evidence for nonlinearity. That is, the Bayes factor comparing the threshold autoregressive model against the linear autoregressive model is roughly 10. In other words, if we had ignored  $M_{TVP}$  we could have been misled into concluding inflation followed a TAR process, when in reality it exhibits structural instability.

Table 3 presents a mixed message and exhibits a pattern we have found common in our Bayesian work with nonlinear time series models. That is, the data are often not informative to definitively decide on a particular class of models. The industrial production data likely exhibits structural instability (61%), but threshold nonlinearity received appreciable support (37%) and even linearity is not ruled out completely (2%).

In this section, we have considered only three different macroeconomic series and a limited class of

<sup>9</sup>In fact, this finding was so strong we were initially suspicious of it. However, an examination of the values of the likelihood function at the MLE indicates the same pattern, i.e. the TVP model with p=4 yields a much higher likelihood function than any of the other models.

Table 3: Posterior Model Probabilities for DIP

	$M_L$	$M_{TAR}$	$M_{TVP}$	Total
p=1	0.004	0.063	$1.4 \times 10^{-4}$	0.067
p=2	0.011	0.062	0.013	0.086
p=3	0.004	0.224	0.094	0.321
p=4	0.001	0.022	0.503	0.525
Total	0.019	0.371	0.609	1

models (e.g. only the simplest two regime TAR). With this in mind, we would argue that our results are cautionary for nonlinear time series econometricians. In previous work (Koop and Potter, 1999,1998a,b), we have argued that classical econometric findings of predictable nonlinearities (e.g. of the sort implied by various regime shift models such as TAR or Markov switching) could be fragile due to data mining. We argued that a Bayesian approach which averages over all models and parameters is less subject to this criticism and tends to indicate weaker support for nonlinearities. The present paper strengthens our previous arguments and extends them to say that some apparently nonlinear time series could in fact merely be structurally unstable. Note that none of our three widely differing series provide strong evidence for the TAR models. However, if we had only compared  $M_L$  and  $M_{TAR}$  we would have concluded that there was strong evidence for the latter class of models for two of the three time series considered. Furthermore, classical econometric analyses have often found the third series, DGDP, to be nonlinear.

One reason for the poor performance of the TAR model is undoubtedly due to the fact that it is very parameter rich and the Bayesian methodology we use here incorporates a strong reward for parsimony. Hence, one message coming out of the present paper is that there might be gains in developing more parsimonious nonlinear models. The endogenous delay threshold autoregressive model of Pesaran and Potter (1997), further developed in Kapetanios (1998), offers promising initial results.

It is instructive to compare the present results to those in Koop and Potter (1998b). The latter paper investigated whether apparent nonlinearities were due to simple structural breaks or outliers and used the DGDP and DIP series. In terms of the conditional mean of these series, there seemed to be reasonable support for linearity (and some support for models with outliers) for DGDP and strong support for the

nonlinear TAR model (and very little support for outlier models) for DIP. The present paper modifies the latter result by arguing that the apparent TAR finding could be due to structural instability in the DIP series.

We have argued in this paper that our prior choice is sensible and that results should not be sensitive to reasonable variations to this choice. Readers who are skeptical of this claim are directed to Appendix B where we use a prior which is much more noninformative than the present ones. Results are qualitatively similar to those presented here.

## 7 Conclusions

In this paper we have done five things:

i) We have recommended a modelling/testing strategy for nonlinear time series. Rather than just comparing linear to standard nonlinear models, we recommend also considering time-varying parameter models. That is, the nonlinear model typically has linear dynamics which differ across two or more regimes in a way that is predictable using past data. This should be thought of as an intermediate case between linear models (linear dynamics do not change at all over time) and time-varying parameter models (linear dynamics change over time in a completely unpredictable way).

ii) We have argued for a methodology based on Bayes factors. In previous work (Koop and Potter, 1999), we have presented many arguments in favor of such an approach. In the current paper, we additionally emphasize the way that Bayes factors allow for the direct comparison of multiple models.

iii) We have shown how to operationalize the general ideas raised in i) and ii). In particular, we discuss useful classes of models, practical techniques for prior elicitation and computational methods.

iv) We have shown that the methods can be implemented in practice in artificial and real data.

v) We have shown that traditional methodologies, which do not consider the structurally unstable model, would have lead us incorrectly to conclude that threshold nonlinearity was present in two of our three series.

## 8 Appendix A: Computational Details

The computational techniques necessary for Bayesian inference and marginal likelihood calculation for the AR and TAR models are standard and will not be repeated here (see Koop and Potter, 1998a,b). For the TVP model, the Gibbs sampler is composed of the following blocks:

$$p(\theta; \mathcal{Y} | \text{Data}; \mathbb{R}^n) = p(\theta | \mathcal{Y}; \text{Data}; \mathbb{R}^n) p(\mathcal{Y} | \text{Data}; \mathbb{R}^n);$$

where the former of these is the Normal density with mean:

$$\mu = (A^{-1} + X^0 X) X^0 y^n$$

and variance:

$$\frac{1}{2} (A^{-1} + X^0 X)^{-1};$$

In the previous expressions,  $y^n$  is a vector with  $t^{\text{th}}$  element  $y_t^n$  (which is defined in Section 4) and  $X$  is a matrix with  $t^{\text{th}}$  row given by  $(1; y_{t-1}; \dots; y_{t-p})$ . The conditional density for  $\beta$  is inverted Gamma with parameters  $\hat{\beta}_{\beta} = T + \hat{\beta}_{\beta}$  and

$$\tau_{\beta}^2 = \frac{\hat{\beta}_{\beta}^{-1/2} + (y^n - X\mu)^0 (y^n - X\mu) + \mu^0 A^{-1} \mu}{\hat{\beta}_{\beta}};$$

The second block in the Gibbs sampler is for  $p(\sigma_i | \text{Data}; \mathbb{R}^n; \theta; \beta)$  which is the product of  $p+1$  inverted Gamma distributions each with parameters  $\hat{\beta}_{\sigma_i} = T + \sigma_i$  and

$$\tau_{\sigma_i}^2 = \frac{1 + \text{SSE}_{\sigma_i}}{\hat{\beta}_{\sigma_i}};$$

where

$$\text{SSE}_{\sigma_i} = \frac{\sum_{t=1}^T (\hat{\theta}_{it} - \hat{\theta}_{it-1})^0 (\hat{\theta}_{it} - \hat{\theta}_{it-1})}{\beta^2};$$

The remaining conditional is for  $\theta^n$  which is Normal. Drawing from this can be done directly using the techniques of de Jong and Shephard (1995). Since these techniques are described in detail in the latter paper, we do not repeat them here.

With this specification involving three blocks, we can directly use the techniques described in section 2.1.2 of Chib (1995) to calculate the marginal likelihood. This requires the evaluation of the prior, likelihood and posterior conditionals at a point in the parameter space. We select the posterior mean (based on an initial run through the Gibbs sampler) as this point.

## 9 Appendix B: Prior Sensitivity Analysis

In this appendix we present results based on a prior which is similar to the one in the body of the paper, but much more noninformative. Loosely speaking, we leave the prior located in roughly the same place but increase the dispersion. In particular, all hyperparameters are the same except we now set  $A=1E1$ ,  $\rho_{\phi} = :01$  and  $\rho_{\psi_i} = :01$  for  $i=0,\dots,p$ . Note that this makes the prior on the AR coefficients very flat, allocating a great deal of weight to explosive regions of the parameter space. The prior degrees of freedom parameters can be interpreted as a prior "sample size". Setting these values to 0.01 implies that our prior has the same weight as one-hundredth of a data point, roughly speaking. In other words, these priors are very noninformative and place a lot of weight in nonsensical areas of the parameter space (e.g. explosive or regions with huge variation in AR coefficients). Nevertheless, as can be seen from Tables B1, B2 and B3, the qualitative results of the paper are not altered through using such a prior.

Table B1: Posterior Model Probabilities for DGDP

	$M_L$	$M_{TAR}$	$M_{TVP}$	Total
p=1	0.748	0.113	0.001	0.862
p=2	0.111	0.013	0.001	0.125
p=3	0.012	$2.9E10^{i 4}$	$3.4E10^{i 5}$	0.012
p=4	0.001	$7.4E10^{i 6}$	$3.5E10^{i 11}$	0.001
Total	0.872	0.127	0.002	1

Table B2: Posterior Model Probabilities for In°ation

	M <sub>L</sub>	M <sub>TAR</sub>	M <sub>TVP</sub>	Total
p=1	3.4E10 <sup>i 9</sup>	2.2E10 <sup>i 9</sup>	1.2E10 <sup>i 8</sup>	1.8E10 <sup>i 8</sup>
p=2	5.2E10 <sup>i 9</sup>	1.8E10 <sup>i 9</sup>	6.4E10 <sup>i 8</sup>	7.1E10 <sup>i 8</sup>
p=3	9.7E10 <sup>i 8</sup>	5.7E10 <sup>i 8</sup>	5.4E10 <sup>i 7</sup>	6.9E10 <sup>i 7</sup>
p=4	2.7E10 <sup>i 6</sup>	4.0E10 <sup>i 4</sup>	1.000	1.000
Total	1.8E10 <sup>i 8</sup>	4.0E10 <sup>i 4</sup>	1.000	

Table B3: Posterior Model Probabilities for DIP

	M <sub>L</sub>	M <sub>TAR</sub>	M <sub>TVP</sub>	Total
p=1	0.019	0.208	5.7E10 <sup>i 5</sup>	0.228
p=2	0.029	0.048	2.3E10 <sup>i 5</sup>	0.077
p=3	0.005	0.041	0.174	0.221
p=4	3.7E10 <sup>i 4</sup>	0.001	0.473	0.474
Total	0.054	0.299	0.647	1

## 10 Bibliography

### References

- [1] Beaudry, P. and Koop, G. (1993). "Do recessions permanently change output?" *Journal of Monetary Economics*, 31, 149-163.
- [2] Chib, S. (1995). "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313-1321.
- [3] Dejong, P. and Shephard, N. (1995). "The simulation smoother for time series models," *Biometrika*, 82, 339-350.
- [4] Greasley, D. and Oxley, L. (1994). "Rehabilitation sustained: The Industrial Revolution as a macro-economic epoch," *Economic History Review*, 2nd Series, 47, 760-768.
- [5] Judge, G., Griffiths, W., Hill, R.C. and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, second edition, New York: John Wiley and Sons.
- [6] Kapetanios, G. (1998). "Essays on the econometric analysis of threshold models," unpublished PhD dissertation, University of Cambridge.
- [7] Koop, G. (1996). "Parameter uncertainty and impulse response analysis," *Journal of Econometrics*, 72, 135-149.

- [8] Koop, G. Pesaran, M.H. and Potter, S.M. (1996). \Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74, 119-148.
- [9] Koop, G. and Potter, S.M. (1999). \ Bayes factors and nonlinearity: evidence from economic time series," *Journal of Econometrics*, 88, 251-85.
- [10] Koop, G. and Potter, S.M. (1998a). \Dynamic asymmetries in US unemployment," *Journal of Business and Economic Statistics*, forthcoming.
- [11] Koop, G. and Potter, S.M. (1998b). "Nonlinearity, structural breaks or outliers in economic time series?" to appear in *Nonlinear Econometric Modelling*, William Barnett et al. (ed.), Cambridge: Cambridge University Press.
- [12] Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Shin, Y. (1992). \Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54, 159-178.
- [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] Min, C. (1992). \Unconditional estimation of time-varying-parameter models: A Gibbs sampling approach," manuscript.
- [15] Pesaran, M.H., and Potter, S. (1997). \A Floor and ceiling model of US output," *Journal of Economic Dynamics and Control* 21, 661-695.
- [16] Poirier, D. (1995). *Intermediate Statistics and Econometrics*, Cambridge: The MIT Press.
- [17] Potter, S. (1995). \A Nonlinear approach to US GNP," *Journal of Applied Econometrics*, 10, 109-125.
- [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.
- [19] Terasvirta, T. and Anderson, H. (1992). \Characterising nonlinearities in business cycles using smooth transition autoregressive models," *Journal of Applied Econometrics*, S119-S136.

Artificial AR(1)-TVP Data Sets with  $\Lambda(0)=\Lambda(1)$

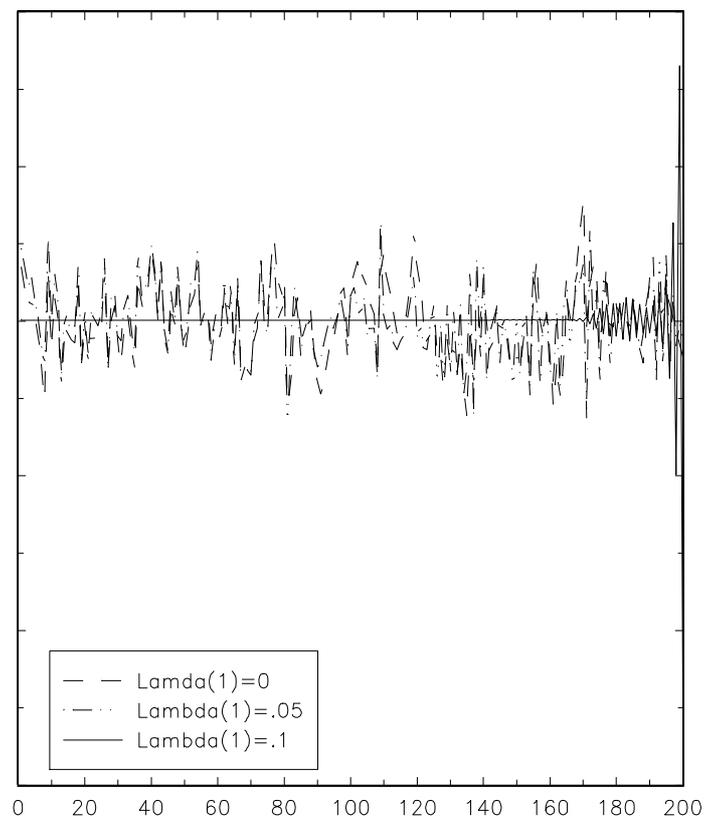


Figure 1:

Artificial AR(1)-TVP Data Sets with  $\text{Lambda}(0)=1$

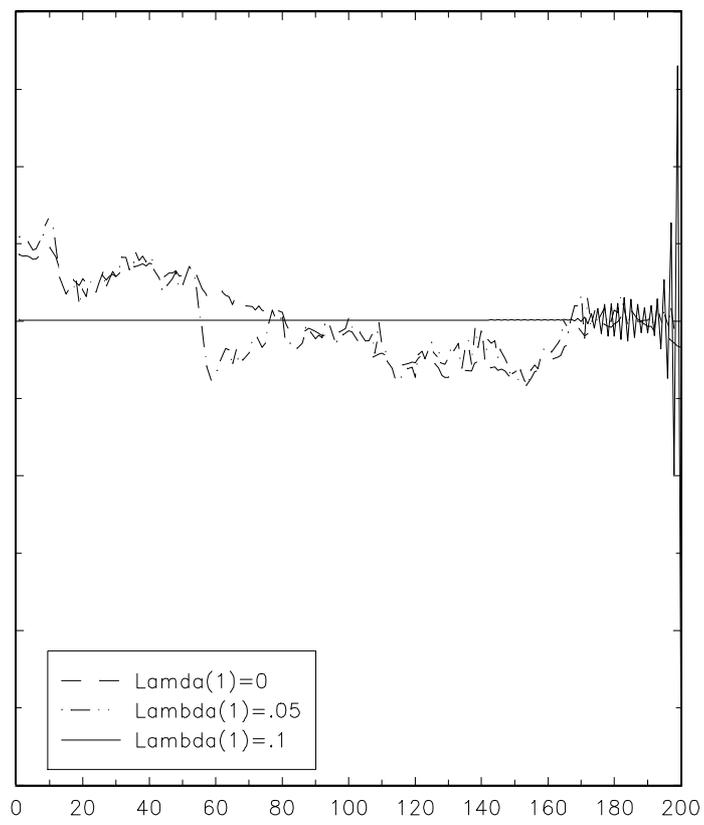


Figure 2:

Experiment 1: Bayes Factors in Favor of Linearity

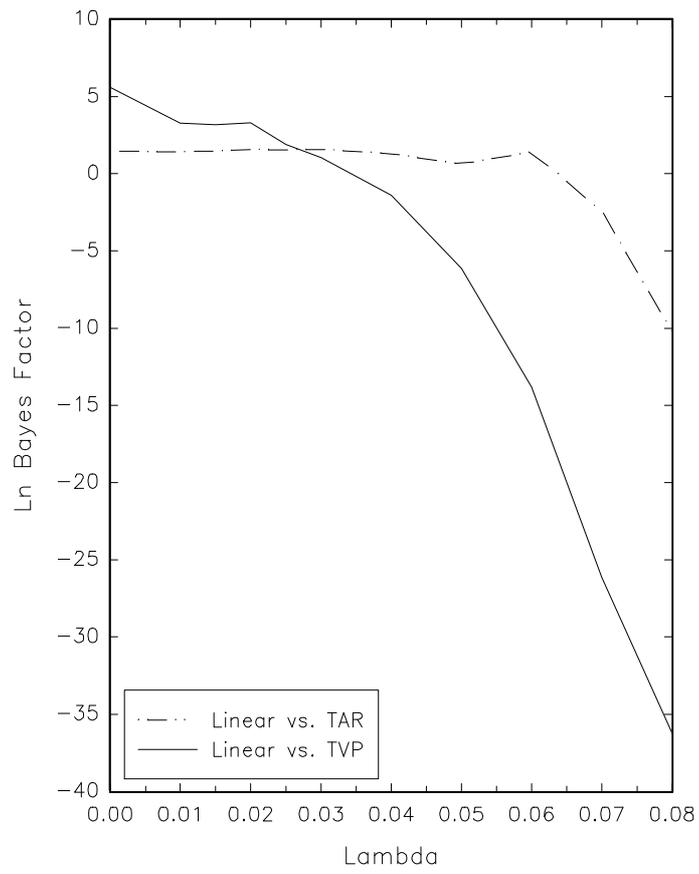


Figure 3:

Experiment 2: Bayes Factors in Favor of Linearity

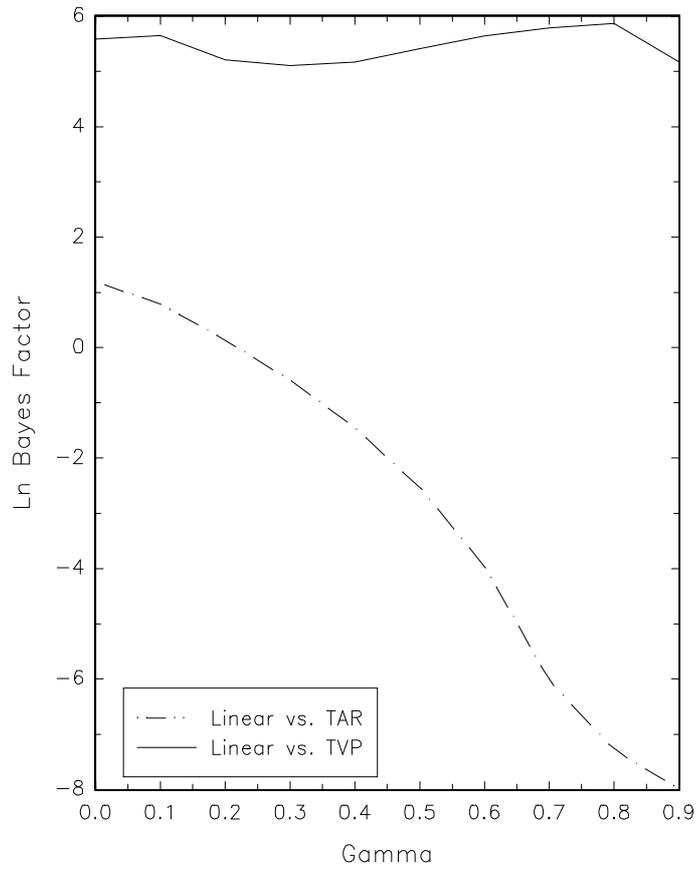


Figure 4:

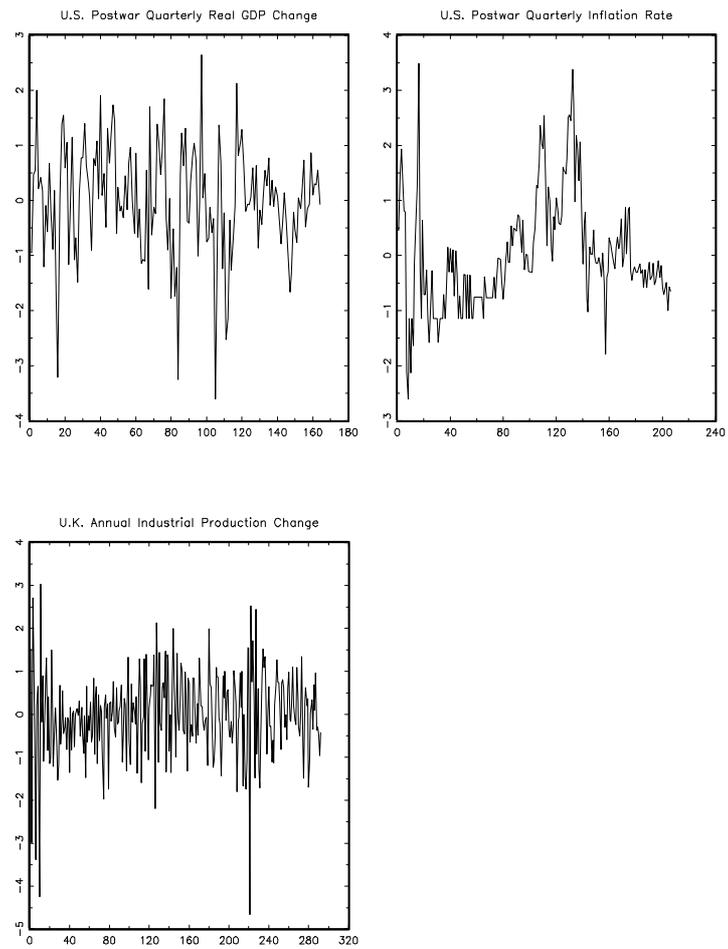


Figure 5: