Learning, Forecasting and Structural Breaks[1]

July 2004

John M. Maheu
Department of Economics
University of Toronto
Toronto, Ontario
M5S 3G7
Canada
jmaheu@chass.utoronto.ca

Stephen Gordon
Département d'économique and CIRPÉE
Université Laval
Quebec City, Quebec
G1K 7P4
Canada
stephen@ecn.ulaval.ca

Abstract

The literature on structural breaks focuses on *ex post* identification of break points that may have occurred in the past. While this question is important, a more challenging problem facing econometricians is to provide forecasts when the data generating process is unstable. The purpose of this paper is to provide a general methodology for forecasting in the presence of model instability. We make no assumptions on the number of break points or the law of motion governing parameter changes. Our approach makes use of Bayesian methods of model comparison and learning in order to provide an optimal predictive density from which forecasts can be derived. Estimates for the posterior probability that a break occurred at a particular point in the sample are generated as a byproduct of our procedure. We discuss the importance of using priors that accurately reflect the econometrician's opinions as to what constitutes a plausible forecast. Several applications to macroeconomic time-series data demonstrate the usefulness of our procedure.

Keywords: Bayesian Model Averaging, Markov Chain Monte Carlo, Real GDP Growth, Phillip's Curve

[1]Both authors are grateful for financial support from the Social Sciences and Humanities Research Council of Canada.

1 Introduction

There is a considerable body of evidence that documents the instability of many important relationships among economic variables. Popular examples include the Phillip's curve, U.S. interest rates and the demand for money.[2] An important challenge facing modern economics is the modeling of these unstable relationships. For econometricians, the difficulties involve estimation, inference and forecasting in the presence of possible structural instability.

Classical approaches to the identification of break points are based on asymptotic theory - see, for example, Ghysels and Hall (1990), Hansen (1992), Andrews (1993), Dufour, Ghysels, and Hall (1994), Ghysels, Guay, and Hall (1997), and Andrews (2003) - are based on frameworks in which both the pre-break and post-break data samples go to infinity.[3] Moreover, tests for multiple breaks such as those proposed by Andrews, Lee, and Ploberger (1996), and Bai and Perron (1998) are also based on similar hypotheses. In applications that make use of available finite data sets, the empirical relevance of such an assumption may be suspect. Furthermore, the statistical properties of a forecasting model based on classical breakpoint tests are far from clear, since such a model would be based on a pre-test estimator.

In contrast, Bayesian approaches to the problem are theoretically simple, are based on finite-sample inference, and typically take the form of a simple model comparison exercise. Examples include Inclán (1994), Chib (1998), Wang and Zivot (2000), Kim, Nelson, and Piger (2004) and Pasaran, Pettenuzzo, and Timmermann (2004) which employ Markov Chain Monte Carlo sampling methods to make posterior inferences regarding the timing of break points in a given sample.

A feature that is common to both strands of the existing literature is the focus on the *ex post* identification of structural breaks that may have occurred in the past. While this question is of course important in itself, economists are more likely to be interested in making forecasts when the the data generating process is unstable. This is the question that we address here. More precisely, we focus on the question of how to make optimal use of past data to forecast out-of-sample while taking into account the possibility that breaks have occurred in the past, as well as the possibility that they may occur in the future.

We present a general methodology for economic decision making based on a model subject to structural breaks. Our analysis uses adaptive learning via Bayes' rule in order to make optimal use of past data to make forecasts. No assumptions concerning break times or their impact on model parameters are made.[4] As a byproduct, our procedure

[2]Stock and Watson (1996) suggest that the laws of motion governing the evolution of many important macroeconomic time series appear to be unstable.

[3]The theory of Dufour, Ghysels, and Hall (1994) and Andrews (2003) only requires that the structurally stable portion of the data goes to infinity.

[4]Pasaran, Pettenuzzo and Timmermann focus on forecasting in the presence of breaks. Extending the work of Chib (1998) and Carlin, Gelfand, and Smith (1992), breaks are assumed to be governed by a first-order Markov chain, while break parameters are drawn from a common distribution. Nonlinear

also provides estimates for the posterior probability that a break point occurred anywhere in the sample, as well as the posterior distribution for the number of observations that would be useful in making out-of-sample forecasts at any point in the sample.

The approach of this paper is most easily seen with a simple example. Consider the observations $y_j, \;\; j = 1, ..., t$ and suppose that an analyst wishes to forecast the value of y_{t+1}. Let M denote the hypothesis that y_{t+1} will be drawn from the same process that generated the previous observations. Conditional on M being true, then $I_t \equiv \{y_1, \ldots, y_t\}$ can be used to to construct a predictive data density, denoted by $p(y_{t+1}|I_t, M)$. Suppose now that the analyst wishes to take into account the possibility that M may not be correct, and that y_{t+1} will be drawn from an entirely different distribution; denoted by M^*, the hypothesis that a structural break will occur between t and $t+1$. If M^* is correct, the past data contained in I_t will be useless in predicting y_{t+1}, so the analyst will be obliged to specify a predictive density of the form $p(y_{t+1}|M^*)$ using his subjective judgment.

Suppose that the analyst attaches a subjective probability λ_{t+1} that M^* is true. Then the predictive density of y_{t+1} is the following mixture:

$$p(y_{t+1}|I_t) = p(y_{t+1}|I_t, M)[1 - \lambda_{t+1}] + p(y_{t+1}|M^*)\lambda_{t+1},$$

from which any quantity of interest, such as predictive moments or quantiles, can be calculated.[5] The first term on the right hand side is the predictive density conditional on no structural break (using all available data) weighted by the probability of no structural break. The second term is the predictive density conditional on a structural break based on the subjective beliefs of the analyst, weighted by the probability of a structural break.

After observing y_{t+1}, the information set is updated so that $I_{t+1} = \{I_t, y_{t+1}\}$. The probability that there was a structural break between t and $t+1$ can be updated according to Bayes' rule:

$$p(M^*|I_{t+1}) = \frac{\lambda_{t+1} p(y_{t+1}|I_t, M^*)}{[1 - \lambda_{t+1}]p(y_{t+1}|I_t, M) + \lambda_{t+1} p(y_{t+1}|I_t, M^*)}$$

and $p(M|I_{t+1}) = 1 - p(M^*|I_{t+1})$. Conditional on M and M^*, the predictive distributions for y_{t+2} can be updated to include the information contained in y_{t+1}. If we denote by $\tilde{M}$ the hypothesis that there is no structural break between $t+1$ and $t+2$, we note that

$$p(y_{t+2}|I_{t+1}, \tilde{M}) = p(y_{t+2}|I_{t+1}, M)p(M|I_{t+1}) + p(y_{t+2}|I_{t+1}, M^*)p(M^*|I_{t+1}).$$

models such as regime switching models (Hamilton (1989)), and time-varying parameter models (Kim and Nelson (1999), Koop and Potter (2001), McCulloch and Tsay (1993) , Nicholls and Pagan (1985)) allow for changes in the form of the conditional density of the data, but these are assumed to evolve according to a fixed law of motion. Parameters change over time, but in a predictable way.

[5]From a Bayesian perspective all inferential information regarding future outcomes of y_{t+1}, conditional on I_t and a model, is contained in the predictive density.

Suppose now that the econometrician wishes to take into account the hypothesis, denoted by $\tilde{M}^*$, that there may be a structural break between $t+1$ and $t+2$. In this case, he would specify a predictive distribution of the form $p(y_{t+2}|\tilde{M}^*)$ and a subjective probability $0 \leq \lambda_{t+2} \leq 1$ that $\tilde{M}^*$ is correct, and the above exercise would be repeated.[6]

At each data point, a new model is introduced and standard results from Bayesian model averaging provide predictive quantities as well as model probabilities. Since a model is defined by the time period it starts, their probabilities can be used to identify breaks in the past. As the empirical examples illustrate, our method provides a rich set of information with regard to model probabilities, posterior parameter densities, and forecasts, all as a function of time. By construction, forecasts incorporate both parameter uncertainty and model uncertainty.

Several applications of the theory to simulated, and macroeconomic time-series data are discussed. In comparison to a model that assumes no breaks, we find the method produces very good out-of-sample forecasts, accurately identifies breaks, and performs well when the data do not contain breaks. In particular, the examples demonstrate that a break can result in a large jump in the predictive variance, which quickly reduces as we learn about the new parameters. However, models that ignore breaks, may have a long-term rise in the predictive variance. As a result, our method can be expected to produce more realistic moments, and quantiles of the predictive density.

We consider predictions of real U.S. GDP and document the reduction in variability discussed in Stock and Watson (2002) among others. Rather than a one time break in volatility, our results point to a gradual reduction in volatility over time with evidence of 3 separate regimes. The model is particularly useful in forecasting the probability of expansions. A second application considers a forecasting model of inflation motivated from a Phillips curve relationship. There are several breaks in this model and a considerable amount of parameter instability. By accounting for these breaks in the process, our approach delivers improved forecasting precision for inflation. The identified breaks, which are not exclusively associated with oil shocks, indicate that the Phillips curve we use is far from stable. However, the structural break model optimally extracts any predictive value from this unstable relationship.

This paper is organized as follows. Section 2 provides a more detailed explanation of the approach, as well as a description of techniques that can be used to implement the procedure. Section 3 applies these methods to simulated data, growth rates in US GDP and to a Phillips curve model of inflation. Section 4 discusses results and extensions.

2 Learning about structural breaks

In this section we provide details of the structural break model and how to forecast in the presence of breaks as well as various posterior quantities that are useful in assessing the impact of structural breaks on a model. In the following we consider a univariate

[6] In the empirical applications we repeat this exercise for all observations in the sample.

time-series context, however, the calculations generalize to multivariate models with weakly exogenous regressors in the obvious way. As in the Introduction, let y_t be an observation, and $I_t = \{y_1, ..., y_t\}$, the information set available at time t.

The first step in making the approach operational is to introduce the role of parameters in learning and forecasting in the presence of structural breaks. Let $p(y_t|\theta, I_{t-1})$ denote the predictive density from a model, conditional on past data I_{t-1}, and conditional on a parameter vector θ. For simplicity, we assume that the form of this density is constant throughout the exercise, and that structural breaks are characterized by unpredictable changes in the value of θ. Extensions to the case where the conditional data density changes over time are straightforward, as noted in Section 2.1 below.

Let M_i be a model in which it is supposed that a structural break occurred between periods $i-1$ and i[7], while $p(\theta|M_i)$ is the prior distribution for θ conditional on the hypothesis that model M_i is correct. In the case of model M_i, data before the break, $y_1, ..., y_{i-1}$ is not useful in parameter estimation, so the posterior $p(\theta|I_{t-1}, M_i)$ is obtained according to

$$p(\theta|I_{t-1}, M_i) = \frac{p(y_i, ..., y_{t-1}|\theta, M_i)p(\theta|M_i)}{\int p(y_i, ..., y_{t-1}|\theta, M_i)p(\theta|M_i)\, d\theta}, \qquad i < t \tag{1}$$

$$p(\theta|I_{t-1}, M_i) = p(\theta|M_i), \qquad i = t. \tag{2}$$

Given the conditional data density $p(y_t|\theta, I_{t-1}, M_i)$ and the posterior $p(\theta|I_{t-1}, M_i)$, the predictive density is

$$p(y_t|I_{t-1}, M_i) = \int p(y_t|\theta, I_{t-1}, M_i)p(\theta|I_{t-1}, M_i)\, d\theta. \tag{3}$$

Since no structure has been put on the form or the rate of arrival of structural breaks, the econometrician is obliged to make a subjective assessment about the probability that a break will occur between $t-1$ and t. This subjective probability is denoted by $0 \leq \lambda_t \leq 1$; λ_t may vary as the econometrician's assessment of the rate of arrival of breaks evolves over time. If $\lambda_j > 0$, $j = 1, .., t-1$ there will be a total of $t-1$ models available at time $t-1$, and whose relative weights will have updated over time according to Bayes' rule. The predictive density for y_t is obtained by integrating across the available models:

$$p(y_t|I_{t-1}) = \left[\sum_{i=1}^{t-1} p(y_t|I_{t-1}, M_i)p(M_i|I_{t-1})\right](1-\lambda_t) + \lambda_t p(y_t|I_{t-1}, M_t). \tag{4}$$

The final term on the right-hand side of (4) is the probability of a break multiplied by the predictive density conditional on a break occurring between $t-1$ and t. Therefore, $p(y_t|I_{t-1}, M_t)$ is based only on the prior $p(\theta|M_t)$, and not a function of past data. However, future data $\{y_{t+1}, y_{t+2}, ...\}$, is used to learn about the new value of θ. If $\lambda_t = 0$, then model M_t receives no weight.

[7] In other words y_{i-1} and y_i are drawn from different data generating processes.

After observing y_t, model probabilities can be updated through Bayes' rule. For instance,

$$p(M_i|I_t) = \frac{(1-\lambda_t)p(y_t|I_{t-1}, M_i)p(M_i|I_{t-1})}{p(y_t|I_{t-1})}, \qquad 1 \leq i < t \tag{5}$$

$$p(M_t|I_t) = \frac{\lambda_t p(y_t|I_{t-1}, M_t)}{p(y_t|I_{t-1})} \qquad i = t. \tag{6}$$

At time t there are a maximum of t models that are being entertained. Any feature of the posterior distribution of θ can be calculated by Bayesian model averaging. For example, if $h(\theta)$ is a function of the parameter vector then its expected value is

$$E[h(\theta)|I_t] = \sum_{i=1}^{t} E[h(\theta)|I_t, M_i]p(M_i|I_t). \tag{7}$$

Similarly, there are $t+1$ models that contribute to the predictive density, and if $g(y_{t+h})$, $h \geq 1$, is a function of y_{t+h} then[8]

$$E[g(y_{t+h})|I_t] = \sum_{i=1}^{t} E[g(y_{t+h})|I_t, M_i]p(M_i|I_t)(1-\lambda_{t+1}) + E[g(y_{t+h})|I_t, M_{t+1}]\lambda_{t+1}. \tag{8}$$

Note that with the appropriate definition of $h(\cdot)$ and $g(\cdot)$, we can recover any moment of interest or probability. Quantiles can be calculated through simulation. In this case, a draw from the models, based on the model probabilities is first taken before a model specific feature is sampled such as a parameter or the simulation of a future observation. Collecting a large number of draws and ordering them allows for the estimation of a quantile.

Since models are identified with particular start-up points, they represent break points. That is, models with high posterior probability identify structural breaks points. For example, if model M_i has a large model probability among all candidate models, it suggests that a break occurred at time i. A plot of the maximum model probability as a function of time may be informative as to when breaks occurs and how confident we are in the new model. Similarly, a time series plot of $E[\theta|I_t]$ presents evidence regarding structural change of the parameter θ through time.

Another useful posterior summary measure of structural breaks is *mean useful observations* (MUO). For instance, if there were 100 data points and it was known that a break occurred in the past at observation 45 then only 55 observations would be useful in estimating the post break model. Mean useful observations is a plot of the expected number of useful observations at a point in time and is calculated at time t as,

$$\text{MUO}_t = \sum_{i=1}^{t} (t-i+1)p(M_i|I_t). \tag{9}$$

[8]Note that we have implicitly assumed one break occurs over the forecast horizon, however, it is possible that multiple breaks occur at $t+1, t+2, ..., t+h$. This can be accounted for, as in Equation (8), by integrating over all possible break permutations.

In the circumstance where no breaks occur, MUO_t as a function of time would be a 45 degree line. However, when a break occurs MUO_t drops below the 45 degree line.

2.1 Prior specification

A coherent model of learning requires that the predictive density for each model M_i must be a proper density. In the current context, where the predictive density $p(y_{t+1}|I_t, M_i)$ defined by the mixture in (3) above, this condition will be satisfied if $p(\theta|I_t, M_i)$ is a proper density. If at time t, the number of observations since the breakpoint associated with model i, that is the difference $t-i$ is sufficiently large, then $p(\theta|I_t, M_i)$ will generally be proper even if the original prior $p(\theta|M_i)$ is an improper 'ignorance prior'.

The use of improper priors or even highly diffuse priors is clearly inappropriate here. It may take many observations before $p(\theta|I_t, M_i)$ is concentrated enough to generate predictive distributions that would receive any significant support. For an econometrician who is trying to generate forecasts using available data, there are significant gains in being able to respond more quickly to the possibility that a break may have occurred recently. Our approach is to adopt proper priors that produce reasonable predictive distributions. Simulating a model based on the prior, and considering the empirical moments is often helpful in selecting prior parameters.

We find it convenient to use the same form for $p(\theta|M_i)$ at each data point in the empirical applications below, but this restriction can be relaxed. For example, if the parameter vector θ can be decomposed according to $\theta = (\theta_0, \theta_1)$, and if the analyst believes that any potential instability can be attributed to changes in θ_0 - that is, the analyst believes that θ_1 is stable - then an appropriate prior might be one where the marginal prior $p(\theta_1|M_i)$ resembles the marginal posterior $p(\theta_1|I_i)$.

Similarly, there is no obvious reason why an analyst should insist on using the same value for the λ_t in every period. If, in his subjective judgment, his model has been producing satisfactory forecasts, and if nothing has occurred that would suggest a recent break, he may choose an extremely low value of λ_t. On the other hand, if he sees a marked decline in the quality of his forecasts, then he might think it appropriate to set λ_t at a larger value. In our inflation example below, we find that an *ad hoc* rule in which λ_t is modeled as an increasing function on the size of past forecast errors does quite well.

The approach discussed in the previous section can also be adapted to the case where the structural instability of the data-generating process is manifested by changes in the form of the conditional data density itself, as noted above. The simplest way would be to simply introduce more than one data density in each period. Suppose that the analyst believes that if there is a break after period t, there are K possibilities for the data density in period $t+1$, denoted by $p^k(y_{t+1}|\theta^k, M_{t+1}^k)$, each accompanied by a prior $p(\theta^k|M_{t+1}^k)$. Suppose also that the analyst assigns the probability λ_{t+1}^k to M_{t+1}^k. In this context, the hypothesis M_{t+1}, that there was a break immediately following period t, is defined by $M_{t+1} \equiv \{M_{t+1}^k\}_{k=1}^K$, and its prior probability is $\lambda_{t+1} = \sum_{k=1}^K \lambda_{t+1}^k$. The predictive distribution conditional on a break, is the mixture

$$p(y_{t+1}|M_{t+1}) = \sum_{k=1}^{K} \left[\int p(y_{t+1}|\theta^k, M_{t+1}^k) p(\theta^k|M_{t+1}^k) \ d\theta^k \right] (\lambda_{t+1}^k/\lambda_{t+1}).$$

After y_{t+1} is observed, the model probabilities $p(M_{t+1}^k|I_{t+1})$ and distributions $p(\theta^k|M_{t+1}^k, I_{t+1})$ are updated in the usual way. Note also that there is no reason to restrict attention to the case where the number of potential models is fixed over time.

2.2 Computational Issues

For many econometric models, all of the quantities discussed in the previous subsection can be calculated using standard Bayesian simulation methods. For an introduction to Markov chain Monte Carlo (MCMC) see Koop (2003) while Chib (2001), Geweke (1998) Robert and Casella (1999) provide a detailed survey of MCMC methods.

The following steps are required for model estimation at time t:

1. Obtain a sample from the posterior of model M_i, $i = 1, ..., t+1$. Calculate posterior quantities for each of the t models or predictive features of interest for the $t+1$ models. Note that M_{t+1} is the prior and an input onto forecasting.

2. Calculate model probabilities, $p(M_i|I_t)$, $\ i = 1, ..., t$.

3. Perform model averaging on quantities of interest, forecasts, etc., using equations (7) and (8). λ_t coupled with the model probabilities from step 2 give the $t+1$ model probabilities used in forecasting.

Typically, and in our applications, we repeat the above steps for all observation $t = 1, ..., T$. However, various schemes in which breaks are permitted at periodic times could be considered by setting the appropriate subset of $\{\lambda_t\}_{t=1}^T$ to 0.

In the special case of the linear model with a conjugate normal-gamma prior an analytical solution is available for the posterior in Step 1. In other cases, Gibbs or Metropolis-Hasting sampling can be used to obtain a sample from the posterior. There are several approaches that can be used to calculate the marginal likelihood and therefore the model probabilities. These include Chib (1995), Gelfand and Dey (1994), Geweke (1995), and Newton and Raftery (1994).

3 Empirical Examples

In this section we present results from simulated data and two examples in macroeconomics. In both cases we consider the following linear model,

$$y_t = X_{t-1}\beta + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{10}$$

where y_t is the variable of interest that is assumed to be related to k regressors X_{t-1} available from the information set I_{t-1}. Independent priors $\beta \sim N(\mu_\beta, V_\beta)$, $\sigma^2 \sim IG(\frac{v_0}{2}, \frac{s_0}{2})$ rule out analytical results, so we use Gibbs sampling to obtain draws from the posterior distribution.[9] If $\theta = [\beta \ \sigma^2]$, then Gibbs sampling produces a set of simulated draws $\{\theta^{(i)}\}_{i=1}^N$ from the posterior distribution after discarding an initial burnin period. In the following examples $N = 5000$, which are collected and the first 100 draws are dropped. To calculate the marginal likelihood of a model and therefore model probabilities we use the method of Geweke (1995) which uses a predictive likelihood decomposition of the marginal likelihood. That is,

$$p(y_1, ..., y_t|M_i) = \prod_{k=1}^{t} p(y_k|I_{k-1}, M_i). \tag{11}$$

Each of the individual terms in (11) can be estimated consistently as

$$p(y_k|M_i) \approx \frac{1}{N} \sum_{i=1}^{N} p(y_k|\theta^{(i)}, I_{k-1}, M_i), \tag{12}$$

which can be conveniently calculated at the end of a MCMC run along with features of the predictive density, such as forecasts.

3.1 Change-points in the mean

As a simple illustration of the theory presented, consider data generated according to the following model,

$$y_t = \mu_1 + \epsilon_t, \quad t < 75 \tag{13}$$
$$y_t = \mu_2 + \epsilon_t, \quad 75 \leq t < 150 \tag{14}$$
$$y_t = \mu_3 + \epsilon_t, \quad t \geq 150 \tag{15}$$

with $\mu_1 = 1$, $\mu_2 = .1$, $\mu_3 = .5$, $\epsilon_t \sim NID(0, .3)$, and $t = 1, ..., 200$. Priors were set to $\mu \sim N(0.2, 9)$, $\sigma^2 \sim IG(25, 10)$, $\lambda_t = 0.01$ for $t = 1, ..., 200$.

Figure 1 displays a number of features of the model predictions. We compare the break model to a nobreak alternative, both with identical priors.[10] Panels A and B show the predictive mean along with the 95% highest density region (HDR) from the predictive density one period out-of-sample. This interval was obtained through simulation from the predictive density based on 5000 draws, which is described as follows. First, a model was randomly chosen based on the model probabilities at time $t - 1$, next a parameter

[9] See Koop (2003) for details on posterior sampling for the linear model with independent but conditionally conjugate priors.

[10] For convenience we label the structural break model as the *break model* and refer to a model that assumes no breaks as the *nobreak model*.

vector was sampled from the posterior simulator and used to simulate the model ahead one observation. The smallest interval from the ordered set of these draws that has 95% probability gives the desired confidence interval

Both sets of confidence intervals are similar before the break at $t = 75$ with the exception of a large positive outlier that the break model briefly interprets as a break. However, after the first break, panel C shows a quick reduction in the predictive mean from the break model while the predictive mean from the nobreak model remains high for a long time. Also note that the confidence intervals for the nobreak model appear to be uncentered relative to the data after the first break.

Similarly, panel D shows the nobreak model to understate the dispersion in the predictive density just after the first break point. On the other hand, the break model correctly identifies a break and consequently has a large increase in the uncertainty about future observations. The second break point is much harder to detect and we only observe a gradual increase in the predictive mean, and predictive standard deviation.

These figures suggest that predictions from the break model should be superior to models that ignore breaks. Table 1 show the improvements in terms of forecasting precision and the log marginal likelihood of both models. We include the root mean squared error (RMSE) and mean absolute error (MAE) based on the predictive mean.[11] Out-of-sample forecasts are included for all observations.[12] Besides the improved forecasts, the estimates for the log marginal likelihoods indicate a log Bayes factor of 22.6 in favour of the break model.

Finally Figure 1E displays the model probabilities through time. This is a 3-dimensional plot of (5) and (6). The model axis displays the models, identified by their starting observation. Note that the number of models is linearly increasing with time. The model probabilities at a point in time can be seen as a perpendicular line from the time axis. At $t = 1$ there is only one model which receives all the probability, at $t = 2$ there are 2 models etc. It can be seen that up until observation 75 model M_1, receives almost probability 1.[13] However, after observing the first break at $t = 75$, model M_{75}, has a probability of .94 which allows the model to quickly adjust to the new data generating process. After this the probability of M_1 drops to zero and M_{75} continues to receive a high probability until the latter part of the sample. The difficulty in detecting the final break at $t = 150$ is clearly seen in this figure with the low hump and dispersed model probabilities in this region.

In additional experiments, we simulated from a structurally stable model (13) for 200 observations. The break model produced very similar results to a no break model. For instance, from one of the simulations, the RMSE for the predictive means was .5622 for the break model and .5604 for the no break model. Other results were very similar across

[11] Based on a quadratic loss function the predictive mean is the Bayes optimal predictor.

[12] The 1st prediction is based only on the prior. Evaluating forecasts based on data in the latter half of the sample, for this example and others, produced the same ranking among models.

[13] From the figure it can just be seen that there is a one time spike in M_{64} at observation $t = 64$ associated with a positive outlier of 2.8456 previously mentioned.

models. This suggests that the approach can be confidently used even when no breaks are present in the data. The next two subsections consider application to forecasting real output and inflation.

3.2 Real Output

A recent literature, beginning with Kim and Nelson (1999), and McConnell and Perez-Quiros (2000), documents a structural break in the volatility of GDP growth (see Stock and Watson (2002) for an extensive review). We consider model estimates and forecasts from an AR(2) in real GDP growth. Let $y_t = 100[\log(q_t/q_{t-1}) - \log(p_t/p_{t-1})]$ where q_t is quarterly U.S. GDP seasonally adjusted and p_t is the GDP price index. Data range from 1947:2 - 2003:3, for a total of 226 observations. The model is

$$y_t = \beta_0 + y_{t-1}\beta_1 + y_{t-2}\beta_2 + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2). \tag{16}$$

Realistic priors were calibrated through simulation[14] and are $\mu_\beta = [.2 \ .2 \ 0]'$, $V_\beta = \text{Diag}(1 \ .03 \ .03)$, $v_0 = 15, s_0 = 10$.[15] We set $\lambda_t = .01$, $t = 1, ..., 226$ which implies an expected duration of 25 years between breaks points The results presented impose stationarity; removing this constraint produces similar results.

Figures 2 - 4 display several features of the estimated model, and Table 2 reports out-of-sample forecasting results. Panels A and B of Figure 2 present the data along with the predictive mean and the associated HPD interval for the break and nobreak AR(2) model, both with the same prior specification. Both models produce very similar predictions as can be seen in panel C of this figure. However, the confidence interval from 1990 onward is noticeably narrower for the break model. The differences in the predictive standard deviations are easily seen in Figure 2D. There is a clear reduction in the standard deviation beginning from the end of the 1980s for the break model, as well as a less pronounced reduction in the 1960s. In contrast the nobreak model estimates the predictive standard deviation as essentially flat with only a slight reduction over time.

The evidence for structural breaks can be seen in Figure 2E. For instance, there is some weak evidence of a break in the 1960s, however as we add more data the probability for a break during this period diminishes. Notice, however, that from the 1960s on, there is always some uncertainty about a break as new models are introduced. This can be seen from the small ridges on the 45 degree line between Models and Time. The final ridge in this plot is associated with a break in 1983:3. This is more clearly seen in Figure 3 which shows the very last line of the model probabilities in Figure 2E based on the full sample of data. There is some uncertainty as to when the break occurs with

[14] When we simulated artificial data using this prior, the 95% confidence regions for the unconditional mean, standard deviation, and the 1st order autocorrelation coefficient are (-2.05,2.71), (0.64,1.27), and (-0.10,.52) respectively.

[15] The priors are conservative, but not unduly so. The proportion of realized observations that lie within the predictive density 95% confidence region, when the posterior always equals the prior, is 0.978.

the maximum probability being associated with model 1983:3. Kim and Nelson (1999), and McConnell and Perez-Quiros (2000) find evidence of a break in 1984:1.[16]

Figures 4A and 4B plot the evolution of the unconditional first and second moment implied by the model. The unconditional mean shows some variability but mostly stays around 1. In other words, the structural breaks do not appear to affect the long-run growth properties of real GDP. In contrast, Figure 4B shows 3 distinct regimes in the unconditional variance. Ignoring the transition periods, the unconditional variance values are 1.8 (1951-1962), 1.2 (1965-1984), and .40 (1990-2003). Rather than a one time break in volatility, our results point to a gradual reduction in volatility over time with evidence of 3 separate regimes.

Table 2 displays out-of-sample results for one-period ahead forecasts. In addition to a quadratic loss function, optimal forecasts are computed for the linear exponential (LINEX) loss function discussed in Zellner (1986). This loss function, $L(y, \hat{y}) = b[\exp(a(\hat{y} - y) - a(\hat{y} - y) - 1)]$ where $\hat{y}$ is the forecast and y is the realized random variable ranks overprediction (underpredictions) more heavily for $a > 0$ ($a < 0$). The table includes $b = 1$ with $a = -1$, and $a = 1$. We report the MAE and RMSE for the predictive mean and for the probability of positive growth next period, $I(y_{t+1} > 0)$, where $I(y_{t+1} > 0) = 1$ if $y_{t+1} > 0$ and otherwise 0.[17] Based on our previous discussion it is not surprising that the MAE or RMSE for both models are very close; neither of these two criteria are affected by the possibility that the predictive variance might be unstable. When the LINEX loss function is used, the break model's ability to capture variations in higher moments provides small gains. On the other hand, the break model produces a 10% reduction in the MAE when forecasting future positive growth as compared to the nobreak model. We also computed longer horizon forecasts (not reported) which provide a similar ranking among the 2 models. Finally, estimates for the log marginal likelihoods indicate a log Bayes factor of 15.6 in favour of the break model.

3.3 Inflation

Another well-known economic relationship that exhibits structural breaks is the Phillips curve, which in its most basic form posits a negative relationship between inflation and unemployment. This structural model has motivated a number of empirical studies that forecast inflation, and many studies have questioned the stability of this relationship (see Stock and Watson (1999) and references therein). It is therefore of interest to see if we can exploit information from this relationship in the presence of model instability.

Define quarterly inflation as $\pi_t = 100 \log(p_t/p_{t-1})$, where p_t is the GDP price index,

[16]There are of course some important methodological differences between their work and our approach. Kim and Nelson (1999) use Bayesian methods but only consider one break while McConnell and Perez-Quiros (2000) is based on the asymptotic theory of Andrews (1993) and Andrews, Lee, and Ploberger (1996).

[17]$E_i I(y_{t+1}) = p(y_{t+1} > 0|I_t)$.

and consider the following model for predicting h-period ahead inflation

$$\pi_{t+h} = \beta_1 + \pi_t\beta_2 + y_t\beta_3 + U_t\beta_4 + \epsilon_{t+h}, \quad \epsilon_{t+h} \sim N(0, \sigma^2) \tag{17}$$

where y_t is the growth rate of real GDP, U_t is the unemployment rate, and $h = 1, 2, 3, 4$. The prior is $\mu_\beta = [.5\ .5\ 0\ 0]'$, $V_\beta = \text{Diag}(1\ .2\ .1\ .1)$, $v_0 = 15$, $s_0 = 5$, $\lambda_t = .01$.[18] Table 3 reports out-of-sample forecasting performance, while Figure 5 displays predictive features of the models, model probabilities, and Figure 6 records the model parameter estimates over time.

Panels A – C of Figure 5 show the predictive mean and associated predictive confidence regions for the break and nobreak models with $h = 4$. As expected, both models clearly lag in responding to inflation, but the break model tends to produce a tighter confidence interval during the 1960s and 1990s. Panel B also shows that the break model does better in adjusting to the increase in inflation during the 1970s following from the oil shock. Note also that the forecasts from the nobreak model in panel A hardly respond during this period. The success of the break model lies in the identification of a break in the process at 1972, as seen in panel D. For instance, based on observations through to 1973, models associated with 1972:1, 1972:2, and 1972:3 receive model probabilities of 0.62, 0.13, and 0.12. Other breaks occur around 1950:2 as a result of the an increase in primary commodity prices and the outbreak of the Korean war, and during the 1981-82 recession that followed the Federal Reserve's decision to target the rate of growth of the monetary base.[19]

The implications for parameter change due to the 3 main breaks that we have identified are found in Figure 6. This figure reports the posterior mean of the break model parameter as a function of time. The most significant changes in the parameters appear to be a temporary increase in the intercept, accompanied by a decrease in the coefficient on unemployment. Estimates for the variance coefficient σ^2 increase during the episodes of high inflation. In addition, there has been a gradual increase in the importance of lagged inflation, β_2, which by the end of the sample achieves a value of 0.5. Finally, note that except for the early part of the sample there is very little evidence that real growth rates or unemployment are important factors in predicting inflation.

Table 3 illustrates the benefits of explicitly dealing with these structural breaks. For each forecast horizon $h = 1, ..., 4$, the structural break model improves on the MAE and RMSE of the nobreak model. In addition, to forecasts of y_{t+h}, we report forecasts of $I(y_{t+h} > 1)$, which on an annual basis is the probability of inflation being in excess of 4%: this may be a useful indicator of high inflation and a quantity of interest to policy makers. As in Tables 1 and 2, we report results for the case where $\lambda_t = 0$ (the nobreak case) and where $\lambda_t = 0.01$, $\forall t$. We also include results where the subjective probability of a break is an increasing function of the standardized forecast error observed in the

[18] This prior specification provides a very conservative predictive density: the proportion of realized observations that lie within the predictive density 95% confidence region, when the posterior always equals the prior is 1, for $h = 1, ..., 4$.

[19] Recall that we are discussing the $h = 4$ case which may affect the identified break point by a year.

previous period, defined by $e_t \equiv (y_t - E[y_t|I_{t-1}])/\sqrt{V[y_t|I_{t-1}]}$. We adopt the following *ad hoc* rule for the choice of λ_t:

$$\lambda_t = [1 + \exp(4.5 - |e_{t-1}|)]^{-1} \quad (18)$$

When $e_{t-1} = 0$, (18) generates $\lambda_t = 0.01$. As e_{t-1} increases in absolute value, so does λ: for $|e_{t-1}| = 1, 2, 3, 4$, we obtain $\lambda_t = 0.03, 0.18$ and 0.38, respectively. We refer to this as the *flexible break* model, while the case in which λ_t is constant is the *fixed break* model.

Given the above discussion and the existing empirical literature on the instability of the Phillips curve, it is probably not surprising that the nobreak model is clearly dominated by the two break models. There are significant gains - in both MAE and RMSE - for using either of the break models to forecast y_{t+h} and $I(y_{t+h} > 1)$ at all forecast horizons considered here. Furthermore, the log Bayes factors - which range between 40 and 80 in favour of the break models - indicate that the data provide overwhelming evidence against the hypothesis that the relationship in (17) is stable over time.

The advantages of using the flexible break model over the fixed break model are less clear. The flexible break model has consistently lower RMSE and MAE when forecasting y_{t+h} at all forecast horizons, but the gains are nowhere near as important as those associated with abandoning the nobreak model. For values of $h = 1, 2$, the fixed break model outperforms the flexible break model in predicting $I(y_{t+h} > 1)$. The log Bayes factors in favour of the flexible break model range from -1.9 to 3.2, so few strong conclusions can be made about the relative merits of the two break models.

4 Discussion

This paper provides a very general approach to dealing with structural breaks for the purpose of model estimation, inference and forecasting: it imposes no structure on the arrival rate of breakpoints, nor does it specify how breaks affect the data generating process. Our empirical examples suggest that being able to quickly identify breakpoints generates significant forecasting efficiency gains in the period immediately following a break. We make a particular emphasis on careful prior elicitation; the focus of interest when specifying priors should be their implications for the predictive distribution of observables. The form of these priors can vary over time, as the analyst learns more non-data-based information. Developing tools for prior elicitation in various forecasting contexts along the lines of the extensions discussed in Section 2.1 will be the subject of future research.

There are also numerical issues that will have to be addressed. In our examples, the number of models is equal to the sample size, but the computational burden is quite modest: computing all the results (including the HPD regions reported for GDP growth rates took just under 25 minutes on a modern Pentium chip based computer. Of course for forecasting in real time, only the models available at time t (in our case, $t+1$ models) need to be estimated at time t, and importance sampling techniques may further reduce these computations by efficiently using past draws from the posterior simulator (Geweke (1995)). In other settings, such as in finance or labor econometrics, the datasets are much larger, and it may be impractical to entertain such a large number of models. In this case, allowing for periodic breaks to occur - for example, at a seasonal frequency - may be a practical alternative.

Table 1: Simulation Example

Model	λ_t	MAE	RMSE	log ML
		y_{t+1}	y_{t+1}	
no break		0.57251	.71365	-219.1259
break	0.01	0.50711	.63119	-196.5148

This table reports mean absolute error (MAE), and root mean squared error (RMSE) for the predictive mean forecast one-period ahead, and the log marginal likelihood estimate. The out-of-sample period is based on all 200 observations.

Table 2: Out-of-Sample Forecasting Performance for US Real Output

Model	λ_t	MAE		RMSE		LINEX		log ML
		y_{t+1}	$I(y_{t+1} > 0)$	y_{t+1}	$I(y_{t+1} > 0)$	a=-1	a=1	
no break		.74514	.29947	1.0187	.3616	.6549	.5478	-321.3516
break	.01	.75791	.26945	1.0276	.3546	.6534	.5316	-305.6682

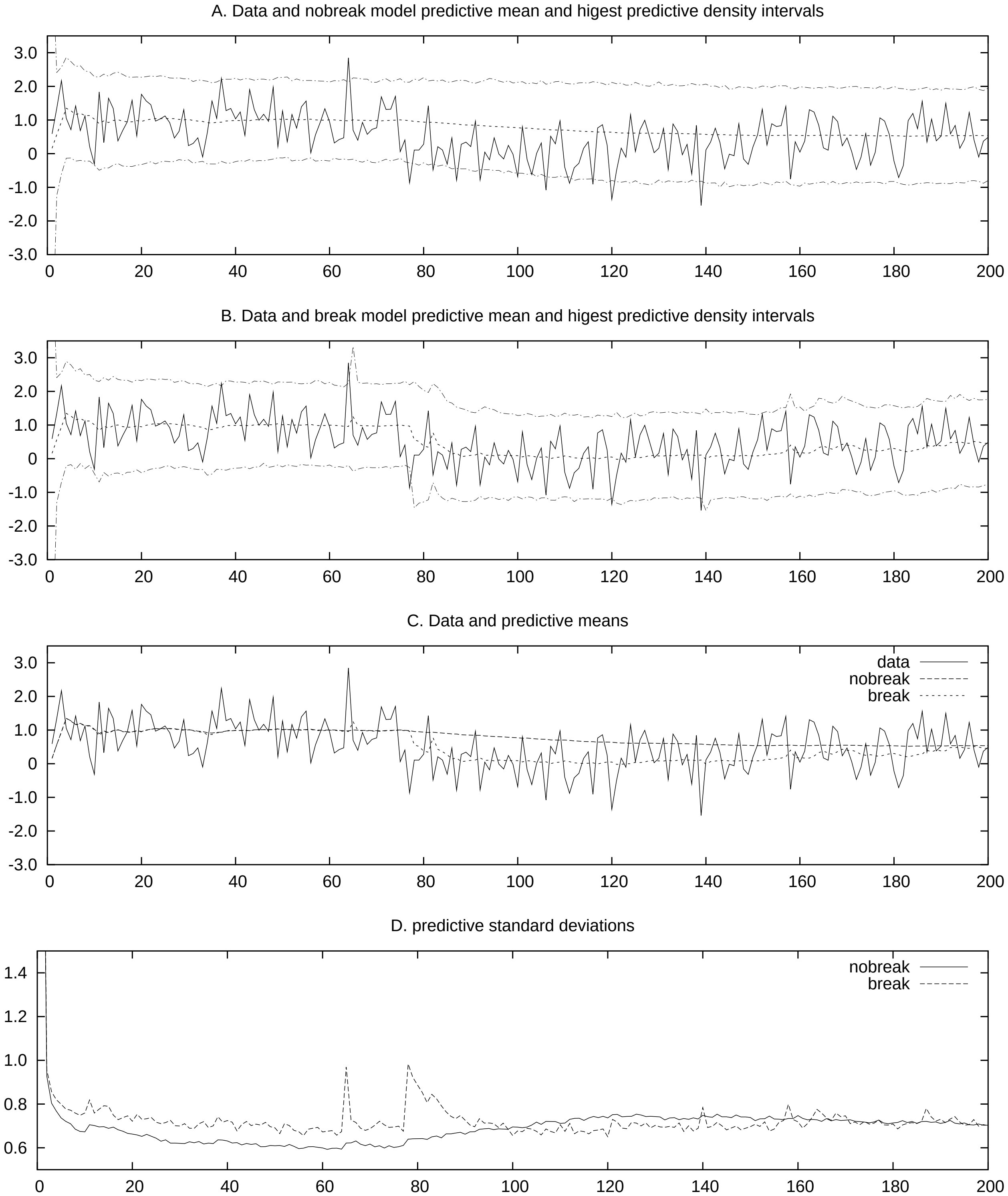
This table reports mean absolute error (MAE), and root mean squared error (RMSE) for the forecasts based on the predictive mean for one-step ahead real GDP growth y_{t+1}, and the positive growth indicator, $I(y_{t+1} > 0)$, where $I(y_{t+1} > 0) = 1$ if $y_{t+1} > 0$ and otherwise 0. In addition average LINEX loss function is reported with $b = 1$, as well as the log marginal likelihood estimate. The out-of-sample period ranges from 1947:4-2003:3 (224 observations).

Table 3: Out-of-Sample Forecasting Performance for Inflation

Model	λ_t	MAE		RMSE		log ML
		y_{t+h}	$I(y_{t+h} > 1)$	y_{t+h}	$I(y_{t+h} > 1)$	
h=1						
no break		.29709	.28514	.4299	.3488	-121.0422
break	.01	.26257	.22854	.4019	.3238	-76.3327
break	$[1 + \exp(4.5 - \lvert e_{t-1} \rvert)]^{-1}$	.26083	.23001	.4016	.3221	-78.2037
h=2						
no break		.36562	.31436	.4931	.3775	-146.3510
break	.01	.30828	.26090	.4414	.3485	-105.4318
break	$[1 + \exp(4.5 - \lvert e_{t-1} \rvert)]^{-1}$	.30360	.26185	.4393	.3463	-106.3143
h=3						
no break		.38128	.32999	.5343	.3882	-164.6276
break	.01	.32086	.24752	.4986	.3403	-108.3239
break	$[1 + \exp(4.5 - \lvert e_{t-1} \rvert)]^{-1}$	.30494	.24349	.4624	.3306	-105.1382
h=4						
no break		.43680	.36440	.5801	.4264	-183.7457
break	.01	.30978	.25864	.4481	.3480	-105.3295
break	$[1 + \exp(4.5 - \lvert e_{t-1} \rvert)]^{-1}$	.30108	.25809	.4384	.3410	-105.1615

This table reports mean absolute error (MAE) and root mean squared error (RMSE) for the forecasts of h-period ahead inflation, y_{t+h}, and a high inflation state indicator $I(y_{t+h} > 1)$, where $I(y_{t+h} > 1) = 1$, and otherwise 0, based on the predictive mean. The out-of-sample period ranges from 1948:2-2003:1. For the break model two cases are presented. In the first, $\lambda_t = .01$ for all data, and the second allows λ_t to be a function of the lagged standardized prediction error e_{t-1}.

Figure 1: Simulated Data



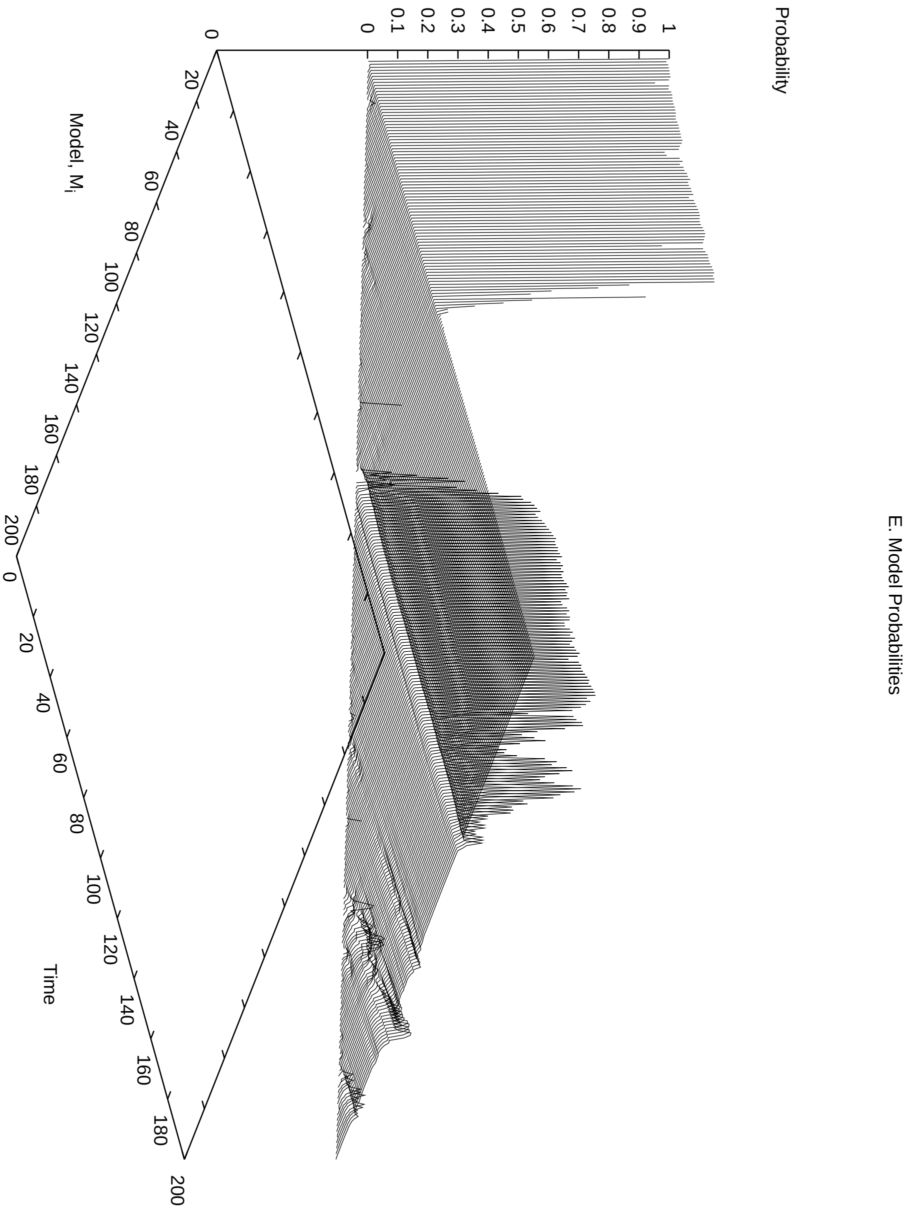
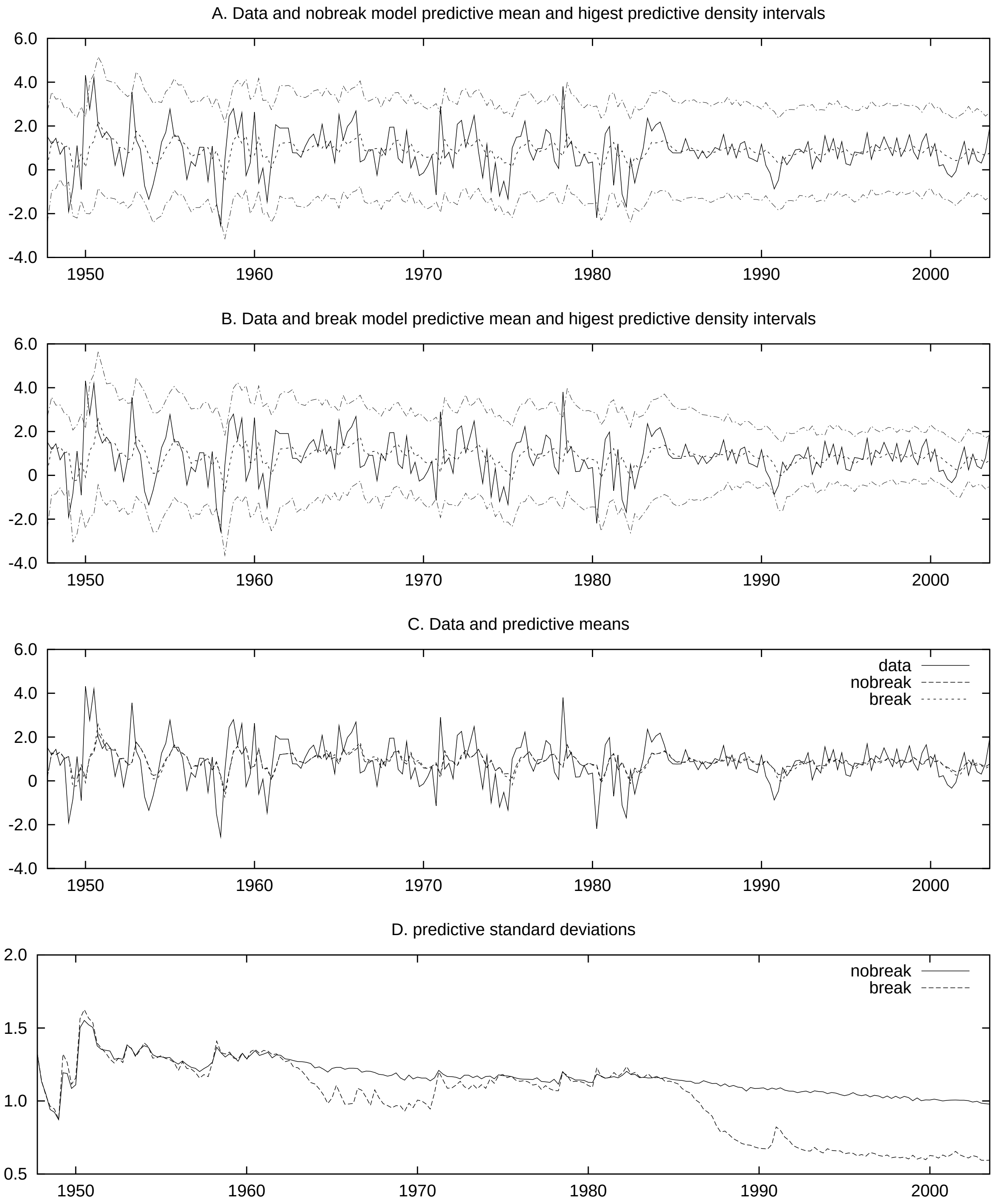

E. Model Probabilities
Probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Model, M_i
0
20
40
60
80
100
120
140
160
180
200
Time
0
20
40
60
80
100
120
140
160
180
200

Figure 2: Real GDP Growth Rates



E. Model Probabilities

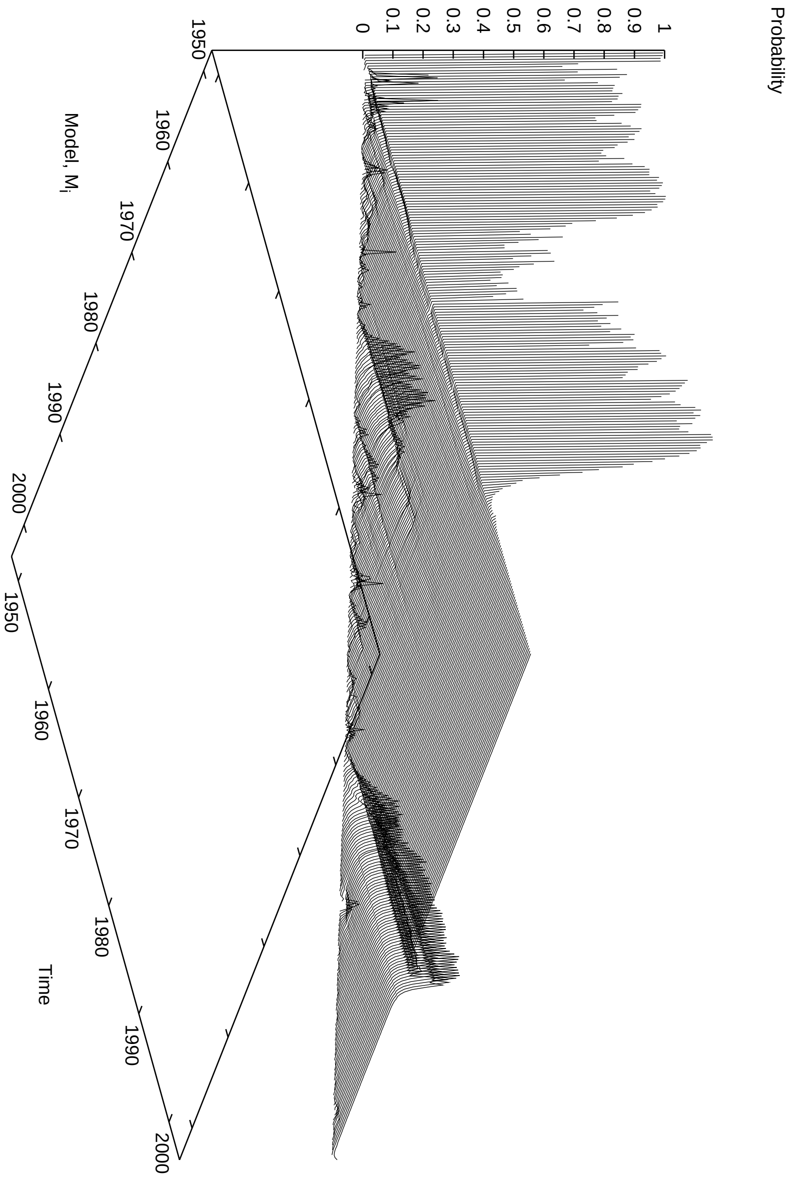


Figure 3: Model Probabilities

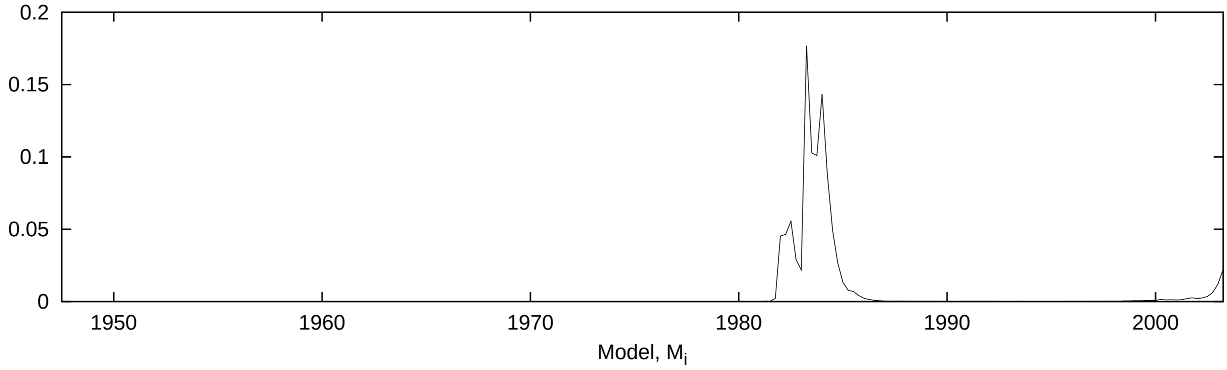


Figure 4: Moments through Time

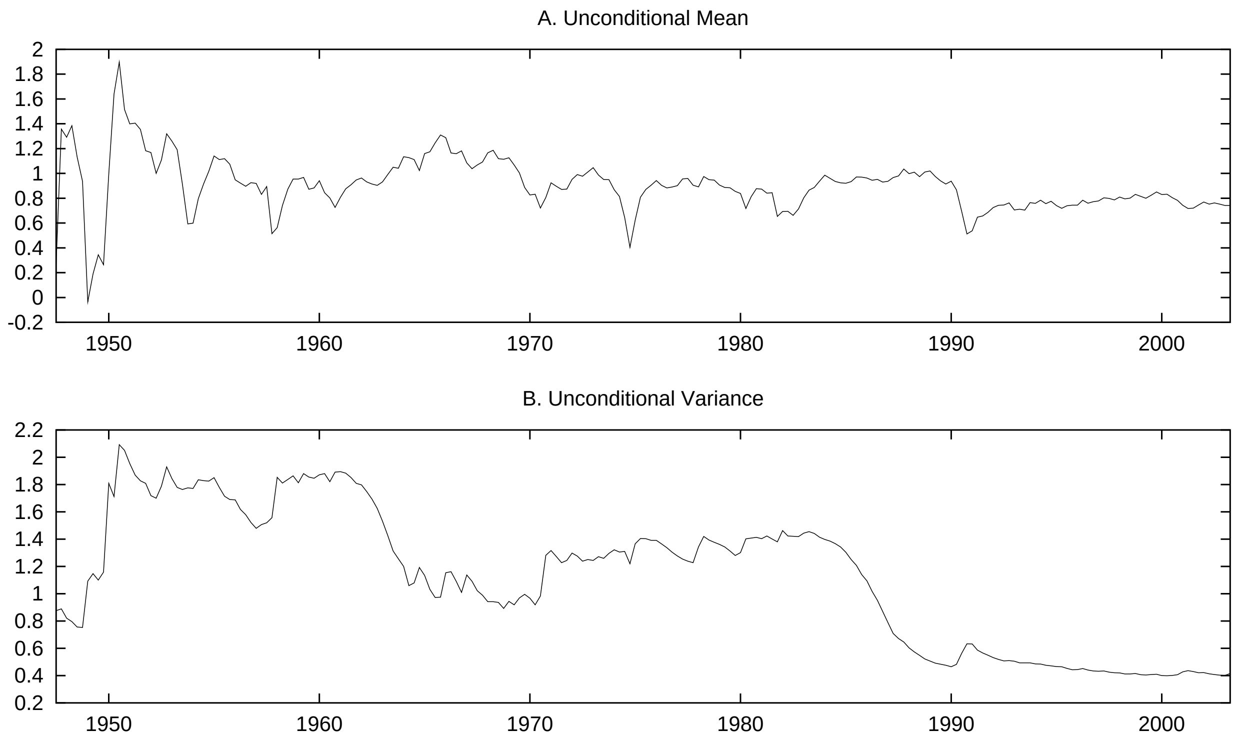
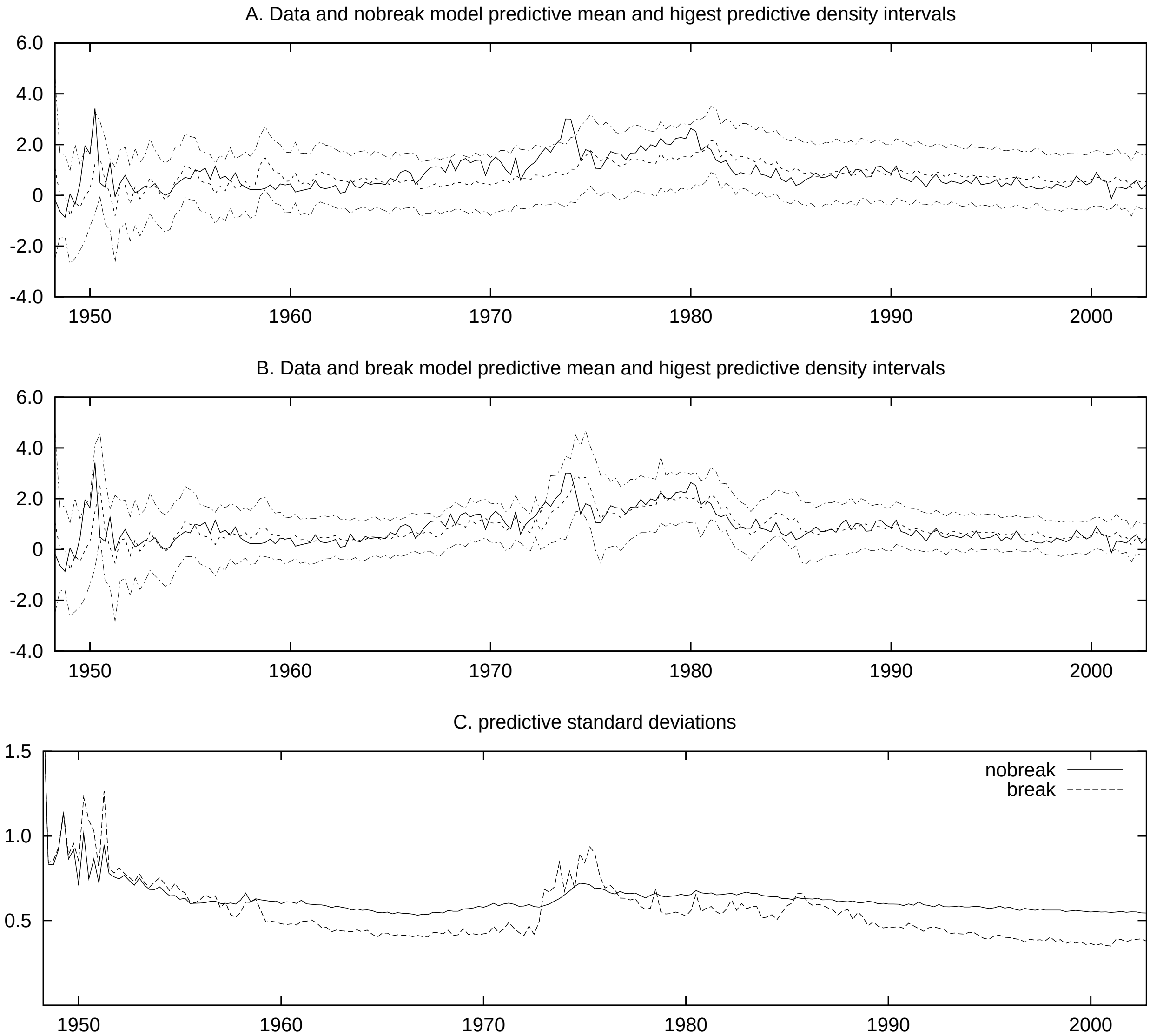


Figure 5: Inflation, h=4



D. Model Probabilities

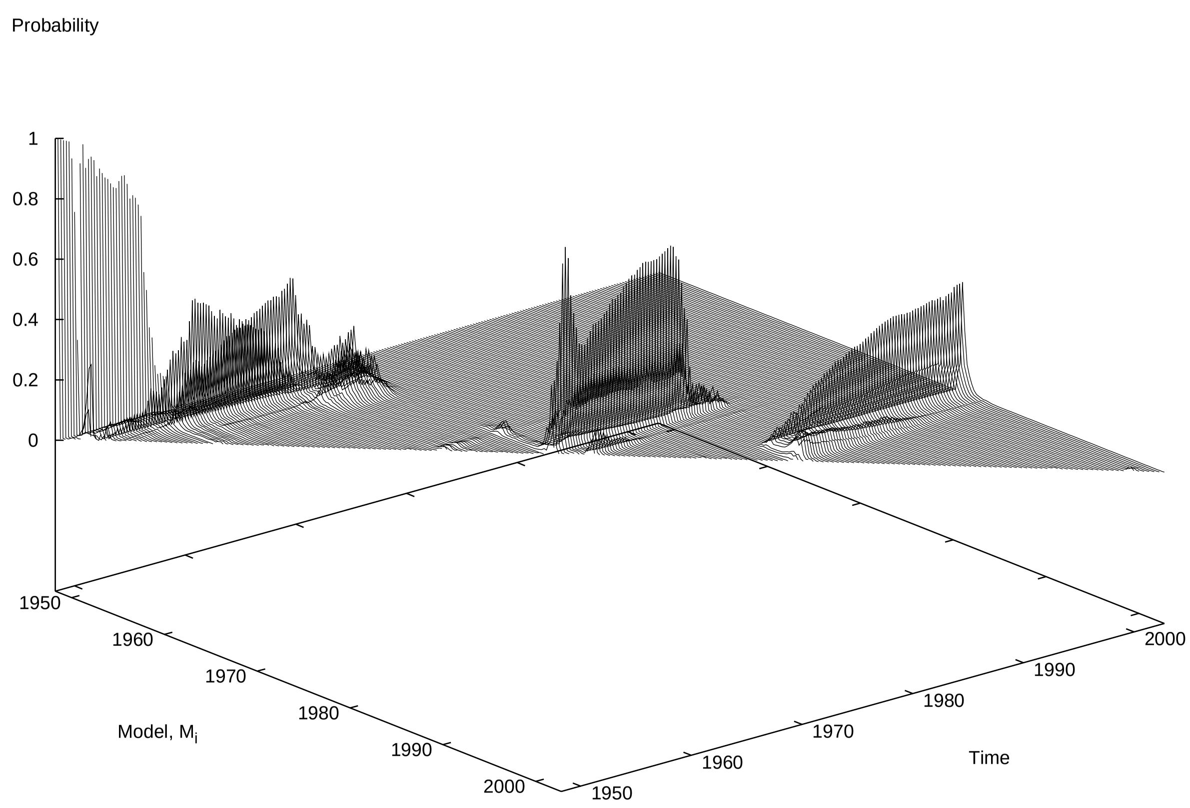
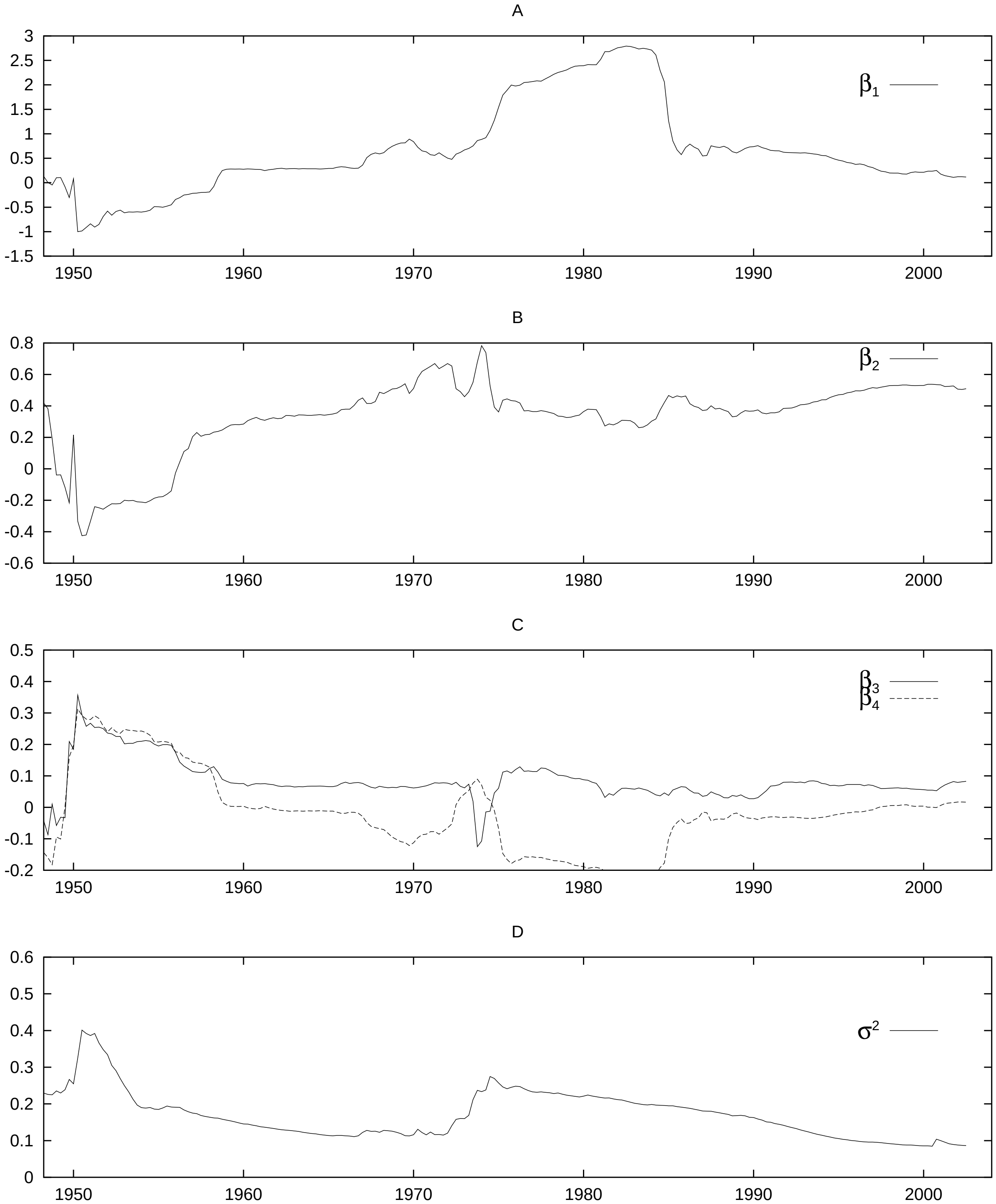


Figure 6: Parameter Estimates through Time, h=4



References

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (2003): "End-of-Sample Instability Tests," *Econometrica*, 71(6), 1661–1694.

Andrews, D. W. K., I. Lee, and W. Ploberger (1996): "Optimal Changepoint Tests for Normal Linear Regression Model," *Journal of Econometrics*, 70, 9–38.

Bai, J., and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Breaks," *Econometrica*, 66(1), 47–78.

Carlin, B., A. E. Gelfand, and A. F. M. Smith (1992): "Hierarchical Bayesian Analysis of Changepoint Problems," *Applied Statistics*, (41), 389–405.

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

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

Chib, S. (2001): "Markov Chain Monte Carlo Methods: Computation and Inference," in *Handbook of Econometrics*, ed. by Heckman, and Leamer. Elsevier Science.

Dufour, J. M., E. Ghysels, and A. Hall (1994): "Generalized Predictive Tests and Structural Change Analysis in Econometrics," *International Economic Review*, 35, 199–229.

Gelfand, A., and D. Dey (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," B, 56, 501–514.

Geweke, J. (1995): "Bayesian Comparison of Econometric Models," University of Iowa.

Geweke, J. (1998): "Comment on Real and Spurious Long-Memory Proporties of Stock Market Data," *Journal of Business & Economic Statistics*, 16(3), 269–271.

Ghysels, E., A. Guay, and A. Hall (1997): "Predictive Tests for Structural Change with Unknown Breakpoint," *Journal of Econometrics*, 82, 209–233.

Ghysels, E., and A. Hall (1990): "A Test for Structural Stability of Euler Conditions Parameters Estimated vis the Generalized Methods of Moments Estimator," *International Economic Review*, 31, 355–364.

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

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business & Economic Statistics*, 10, 321–336.

Inclán, C. (1994): "Detection of Multiple Changes of Variance using Posterior Odds," *Journal of Business & Economic Statistics*, 11(3), 289–300.

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

Kim, C. J., C. R. Nelson, and J. Piger (2004): "The Less-Volatile U.S. Economy" A Bayesian Investigation of Timing, Breath, and Potential Explanations," *Journal of Business & Economic Statistics*, 22(1), 80–93.

Koop, G. (2003): *Bayesian Econometrics*. Wiley, Chichester, England.

Koop, G., and S. Potter (2001): "Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series?," *Econometrics Journal*, 4, 37–55.

McConnell, M. M., and G. P. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?," *American Economic Review*, 90(1464-1476).

McCulloch, R. E., and R. Tsay (1993): "Bayesian Inference and Prediction for Mean and Variance Shilts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.

Newton, M. A., and A. Raftery (1994): "Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion).," B, 56, 3–48.

Nicholls, D. F., and A. R. Pagan (1985): "Varying Coefficient Regression," in *Handbook of Statistics*, ed. by Hannan, Krishnaiah, and Rao. North Holland.

Pasaran, M. H., D. Pettenuzzo, and A. Timmermann (2004): "Forecasting Time Series Subject to Multiple Structural Breaks," Cambridge Working Papers in Economics, 433.

Robert, C. P., and G. Casella (1999): *Monte Carlo Statistical Methods*. Springer, New York.

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

——— (2002): "Has the Business Cycle Changed and Why?," *NBER Macroeconomics Annual*, pp. 159–218.

Stock, J. H., and M. W. Watson (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

Wang, J., and E. Zivot (2000): "A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance," *Journal of Business & Economic Statistics*, 18(3), 374–386.

Zellner, A. (1986): "Bayesian Estimation and Prediction using Asymmetric Loss Functions," *Journal of the American Statistical Association*, 81, 446–451.