scottish institute for research in economics

SIRE DISCUSSION PAPER

SIRE-DP-2013-34

Model Switching and Model Averaging in Time-Varying Parameter Regression Models

Miguel Belmonte
Gary Koop

University of Strathclyde

www.sire.ac.uk

Model Switching and Model Averaging in Time-Varying Parameter Regression Models[*]

Miguel Belmonte

University of Strathclyde

Gary Koop

University of Strathclyde

January 2013

Abstract: This paper investigates the usefulness of switching Gaussian state space models as a tool for implementing dynamic model selecting (DMS) or averaging (DMA) in time-varying parameter regression models. DMS methods allow for model switching, where a different model can be chosen at each point in time. Thus, they allow for the explanatory variables in the time-varying parameter regression model to change over time. DMA will carry out model averaging in a time-varying manner. We compare our exact approach to DMA/DMS to a popular existing procedure which relies on the use of forgetting factor approximations. In an application, we use DMS to select different predictors in an inflation forecasting application. We also compare different ways of implementing DMA/DMS and investigate whether they lead to similar results.

Keywords: Model switching, forecast combination, switching state space model, inflation forecasting

1 Introduction

Bayesian model averaging or model selection (BMA or BMS) methods are commonly used when the researcher is faced with many models. See, for instance, Hoeting, Madigan, Raftery and Volinsky (1999) and Chipman, George and McCulloch (2001) for surveys of these methods. Numerous empirical applications use these methods. However, they were developed for regression models or other models where parameters are constant over time. In time series econometrics, motivated by strong empirical evidence of structural breaks or other forms of parameter change in many economic variables, models where parameters change over time have long been used. Models such as the time-varying parameter (TVP) regression model have enjoyed great popularity, particularly in macroeconomics [see, among many others, Cogley and Sargent (2005), Cogley, Morozov and Sargent (2005), Primiceri (2005), Koop, Leon-Gonzalez and Strachan (2009), D'Agostino, Gambetti and Giannone (2011) and Korobilis (2012)]. Just as with constant coefficient models, it is possible that the researcher working with TVP regression models will want to do model averaging and selection. However, it will typically be desirable to do these in a time varying manner. This leads to an interest in dynamic model averaging (DMA) or dynamic model selection (DMS). With DMA, the weights used in the model averaging procedure can change over time. With DMS, the model selected can change over time. This distinguishes it from conventional model selection methods where one model is selected and assumed to hold at all points in time.

The literature on DMA or DMS is much more limited than that on BMA or BMS. Perhaps the most prominent DMA approach for use with TVP regression models is that of Raftery, Karny and Ettler (2010). To explain what this algorithm involves, we begin by defining the set of models under consideration. Let y_t be a dependent variable and Z_t be a row vector containing explanatory variables. We have K models which are characterized by having different subsets of Z_t as explanatory variables. Denoting these by $Z_t^{(k)}$ for $k = 1, .., K$, a set of TVP regression models can be written as:

$$
\begin{aligned}
y_t &= Z_t^{(k)}\theta_t^{(k)} + \varepsilon_t^{(k)} \\
\theta_{t+1}^{(k)} &= \theta_t^{(k)} + \eta_t^{(k)},
\end{aligned}
\tag{1}
$$

$\varepsilon_t^{(k)}$ is $N\left(0, \sigma_\varepsilon^{2(k)}\right)$ and $\eta_t^{(k)}$ is $N\left(0, \Sigma_\eta^{(k)}\right)$.[1]

---

[1] Note that we have written the error variances, $\sigma_\varepsilon^{2(k)}$ and $\Sigma_\eta^{(k)}$, as being constant. In empirical work it may be desirable to have one or both of them to be time-varying. Details of our treatment of this issue will be given below.

DMA and DMS can be done by calculating $\Pr\left(s_t = k|y^{t-1}\right)$ for $k = 1, .., K$ where $s_t \in \{1, 2, .., K\}$ denotes which model applies at each time period and $y^t = (y_1, .., y_t)'$. DMS involves selecting, at time t, the single model with the highest value for $\Pr\left(s_t = k|y^{t-1}\right)$. DMA involves averaging across models using these probabilities. Different approaches to DMA or DMS arise when different models or methods are used to calculate $\Pr\left(s_t = k|y^{t-1}\right)$. Raftery et al (2010), working in an application involving many potential explanatory variables and, hence, a large model space, uses forgetting factor methods to approximate $\Pr\left(s_t = k|y^{t-1}\right)$. This leads to a computationally simple algorithm which does not require the use of Markov chain Monte Carlo (MCMC) methods. In applications with many potential explanatory variables [e.g. Raftery et al (2010), Koop and Korobilis (2012) and Koop and Tole (2012)], the algorithm of Raftery et al (2010) does seem to be the only computationally feasible algorithm currently available. However, as discussed in Section 3 of Raftery et al (2010), it is an approximate method that does not arise from a particular statistical model of model switching. Furthermore, it is a filtering algorithm as opposed to a smoothing algorithm. That is, it provides the user with $\Pr\left(s_t = k|y^{t-1}\right)$ for $t = 1, .., T$ as opposed to $\Pr\left(s_t = k|y^T\right)$.

The purpose of this paper is to investigate the use of an alternative, model-based, way of allowing for time-varying model switching and compare it to the algorithm of Raftery et al (2010). This alternative is the family of switching Gaussian state space models described in, among other places, Kim (1994), Kim and Nelson (1999) and Fruhwirth-Schnatter (2001a, b). Switching Gaussian state space models will be described in the following section. Here we note only that they have been occasionally used in econometric applications [see Chapter 13 of Fruhwirth-Schnatter (2006) for a list of applications], but typically for state space models where the system matrices vary across regimes, not for selecting explanatory variables in TVP regression models [an exception being Chan et al (2012)]. An advantage of the use of switching Gaussian state space models is that results are not approximate, being based on a valid Bayesian posterior distribution. A further advantage is that either filtered or smoothed estimates can be obtained using existing algorithms.

A disadvantage of the use of switching Gaussian state space models is that MCMC methods are required. This substantially raises the computational burden and means their usage is limited to model spaces based on relatively few explanatory variables. However, it provides a setting in which we can compare DMA using the algorithm of Raftery et al (2010) to DMA using switching linear Gaussian state space models. If we find the algorithm of Raftery et al (2010) to provide results which are quite different from those using switching Gaussian state space models in a setting with a small model space, it will

raise concerns about the use of Raftery et al (2010)'s algorithm in the large model spaces where it is typically used. However, if the two approaches yield similar results, it will increase our confidence in the use of the algorithm of Raftery et al (2010) in large model spaces.

This paper contains an application involving selecting between or averaging different independently produced forecasts of a dependent variable. That is, Z_t will contain various forecasts of the dependent variable y_t. Methods for combining forecasts provided by different models goes back to Bates and Granger (1969) and Granger (2006) provides a recent survey. Recent approaches related to our own include Guidolin and Timmermann (2009), which uses a Markov switching approach to model switching in constant coefficient models and Billio, Casarin, Ravazzolo and van Dijk (2011) who develop an approach with time-varying forecast weights. Our application is to forecasting US inflation. Papers such as Ang, Bekaert and Wei (2007) consider various forecasts of inflation (e.g. forecasts produced by professional forecasters, consumer surveys, econometric forecasts, etc.) and investigate which ones forecast best. Ang, Bekaert and Wei (2007) find that surveys do. We add to this literature using DMS and DMA methods. Note that, unlike Ang, Beckaert and Wei (2007), we can have forecast switching so that, e.g., consumer surveys forecast best at some points in time and econometric models forecast best at other times. We find that there is evidence of model switching which would be missed by conventional approaches. Our empirical application also provides evidence that the algorithm of Raftery et al (2010) is a reasonable one which yields results which are similar to those provided by the switching Gaussian state space model.

The remainder of this paper is organized as follows. The second section describes how switching Gaussian state space models can be used to do DMS or DMA. The third section describes our application. It is divided into sub-sections which: i) discuss some general issues in combining inflation forecasts from various sources, ii) describe the data, iii) present empirical results using the switching Gaussian state space approach and iv) compare the latter approach to DMA and DMS using the methods of Raftery et al (2010).

2 DMA and DMS Using Switching Linear Gaussian State Space Models

The framework given in (1) is closely related to the switching linear Gaussian state space model discussed, e.g., in Fruhwirth-Schnatter (2006, pages 393-394 and 406-410) who provides several citations, mostly from the engineering literature, of papers which have

used such models. A switching linear Gaussian state space model can be written as:

$$
\begin{aligned}
y_t &= H_t^{[s_t]}\theta_t + \varepsilon_t \\
\theta_t &= F_t^{[s_t]}\theta_{t-1} + \eta_t
\end{aligned}
$$

where y_t is observed, ε_t is $N\left(0, \sigma_\varepsilon^{2[s_t]}\right)$ and η_t is $N\left(0, \Sigma_\eta\right)$. The errors are independent of each other and at all leads and lags. $s_t \in \{1, .., K\}$ follows a Markov switching specification, i.e. we have a Markov transition matrix with elements $\zeta_{ij} = \Pr\left(s_t = i | s_{t-1} = j\right)$ for $i, j = 1, .., K$.

We adapt this specification for use with variable selection in TVP regression models by using particular forms for the system matrices. In particular, we set

$$
\begin{aligned}
H_t^{[s_t]} &= Z_t G^{[s_t]} \\
F_t^{[s_t]} &= I.
\end{aligned}
\tag{2}
$$

In most of our empirical work, we set $Z_t = (z_{1t}, .., z_{Kt})$ to contain K explanatory variables and define $G^{[s_t=k]}$ to be the $K \times K$ matrix which selects the kth explanatory variable. That is, $G^{[s_t=k]}$ is a matrix of zeros except for the $(k, k)^{th}$ element which is set to one. In the final subsection of our empirical work, we consider TVP regression models with more than one explanatory variable and $G^{[s_t=k]}$ is defined to pick out the appropriate sets of explanatory variables.

Defined in this way, $\theta_t = (\theta_{1t}, .., \theta_{kt})'$ is a vector of time-varying regression coefficients. The choice $F_t^{[s_t]} = I$ leads to the conventional choice of random walk evolution of these coefficients. We also let Σ_η be a diagonal matrix with k^{th} diagonal element $\sigma_{\eta k}^2$ so that the regression coefficients evolve independently of one another.

In the main part of our empirical work, Z_t will contain different forecasts of inflation. It can be seen that (2) implies that, when $s_t = k$, the TVP regression model using the kth explanatory variable is used. That is, our model space is composed of K models each containing one explanatory variable. However, we also present some results where we consider TVP regression models with more than one explanatory variable. We allow for all combinations of the K explanatory variables, leading to a model space containing $2^K - 1$ models.

Switching between different TVP regression models is controlled through a Markov switching process with switching probabilities given by ζ_{ij}. Thus, the switching Gaussian

state space model, with system matrices defined as in (2), can be used to do DMS or DMA in the context of single statistical model. And Bayesian methods for posterior inference (filtering and smoothing) in this model are developed in several places, including Fruhwirth-Schnatter (2001a, b). In this paper, we use the algorithm of Fruhwirth-Schnatter (2001a, b) (see the Technical Appendix for details).

3   Application: Selecting the Best Inflation Forecasts

3.1   Introduction

The literature on forecasting inflation is voluminous [see, e.g., Faust and Wright (2012) for a recent survey]. We aim to contribute to the literature on choosing between multiple forecasts of inflation. In an influential paper, Ang, Bekaert and Wei (2007) compare various methods for forecasting inflation including surveys (of professional forecasters and of the public at large) and simple time series forecasting methods. Their main conclusion about which methods forecast best is pithily summarized in the first two words of their abstract: "Surveys do!". Faust and Wright (2012) come to a similar conclusion using different econometric methods. The purpose of our application is to investigate whether this conclusion holds in the context of a more formal statistical modelling procedure involving DMA and DMS. Most importantly, our framework allows us to investigate whether the best forecasting model changes over time. After all, it is possible that the time series econometrician (whose forecasts are based on past patterns in the data) will forecast well in normal times, but forecast poorly around times of changes such as business cycle turning points. Professional economists, who can use qualitative events observed in real time (e.g. the collapse of Lehman Brothers) to aid in their forecasting, may be better forecasters at turning points. DMS and DMA can directly find patterns such as these where the best forecasting procedure changes over time or over the business cycle. Conventional methods, which just aim to find one best forecast procedure, cannot.

3.2   Data

Care must be taken with variable definitions and timing to make sure the forecasts made by forecasters are matched up with the outcomes they are compared to. Given the influence of the paper by Ang, Bekaert and Wei (2007), we follow their choices where possible. The interested reader is referred to Ang, Bekaert and Wei (2007) who discuss the relevant issues in detail. As a timing convention, note that all the t subscripts used below are for the times that the forecasts are being made. So, for instance, in 1996Q1 surveys were

taken about inflation over the upcoming year through $1997Q1$. These are dated as $t = 1996Q1$ in the equations below.

Our dependent variable is CPI inflation. Given that inflation forecasts are typically one-year ahead, we use as our dependent variable an annual inflation rate. To be precise, our dependent variable, π_t^R, is the realized value for inflation over the subsequent year defined as

$$\pi_t^R = \pi_{t+1} + .. + \pi_{t+4},$$

where

$$\pi_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

and P_t is the CPI (Consumer Price Index for All Urban Consumers).

We use four different forecasts of annual inflation rates which can be thought of as coming from four different sets of agents: i) the professional forecasters, ii) consumers, iii) time series econometricians and iv) a naive agent.

The professionals' forecasts of inflation are taken from the Survey of Professional Forecasters (SPF) available through the Federal Reserve Bank of Philadelphia website. Detailed explanation about this data source are also available on this website. The inflation forecast we use, π_t^{SPF}, is the median of the one-year ahead inflation forecasts provided by the professionals.

Consumers' forecasts of inflation are taken from the University of Michigan consumer survey. Surveyed individuals are asked by how much they expect prices to change over the next 12 months. The inflation forecast we use, π_t^{CS}, is the median of their forecasts.

There are dozens of different forecasts of inflation produced by time series econometricians. However, it has proved difficult to beat simple forecasting models by much. For instance, Stock and Watson (2010) argue that it is "exceedingly difficult to improve systematically on simple univariate forecasting models". In this spirit, to represent the time series econometrician, we use an autoregressive model. To be specific, π_t^{TS} is the forecast of the time series econometrician using OLS forecasts from an AR(1) model. Forecasts made at time t are made using information available up to and including time $t-1$. Given that π_t^R is an average over four quarters, this means the model used for these forecasts is:

$$\pi_t^R = \alpha + \rho \pi_{t-4}^R + \varepsilon_t.$$

Finally, we have our naive agent producing simple no-change forecasts, π_t^{NOC}, where the forecaster simply uses the most recently available annual inflation rate as a forecast

for next year's inflation. Thus,

$$\pi_t^{NOC} = \pi_{t-1} + .. + \pi_{t-4}.$$

We stress that π_t^{TS} and π_t^{NOC} will be forecasts made at time t of inflation one year in the future.

All data except π_t^{SPF} is taken from the Federal Reserve Bank of St.. Louis' FRED database.[2] Our forecasts runs from $1981Q3$ through $2011Q2$ (i.e. the last forecast is made in $2011Q2$ which can be compared the the actual inflation outcome through $2012Q2$). Figure 1 plots the data.
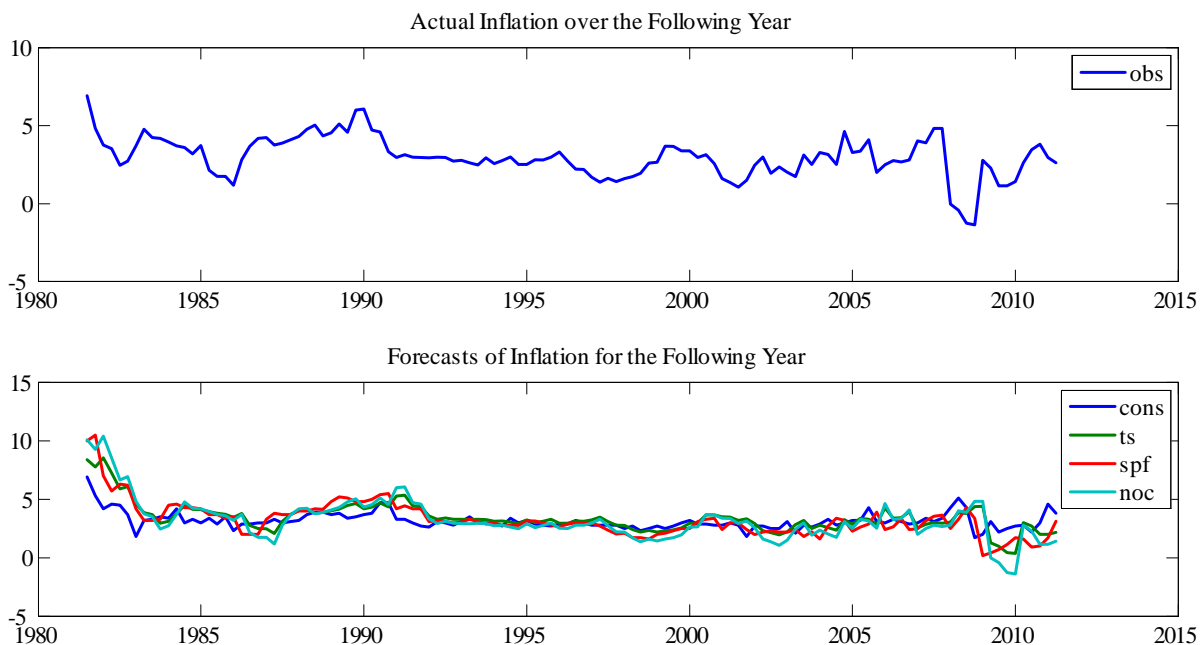

Figure 1

In terms of the notation used in Section 2, $y_t = \pi_t^R$ and $Z_t = \left(\pi_t^{SPF}, \pi_t^{CS}, \pi_t^{TS}, \pi_t^{NOC} \right)$.

3.3   Which Inflation Forecasts are Best?

Our main interest is in which of the four forecasts has been best at each point in time. To shed light on this, Figure 2 presents smoothed estimates of the probabilities,

---

[2]Where relevant, monthly data has been made into quarterly data by taking the observation for the last month of the quarter. See Ang, Bekaert and Wei (2007), page 1171.

$\Pr\left(s_t = k|y^T\right)$ for $k = 1,..,4$ and $t = 1,..,T$ using the switching Gaussian state space model.

Our results have some similarities with the general findings of Ang, Bekaert and Wei (2007) and Faust and Wright (2012). Consumer surveys, in particular, do tend to be chosen as the best model in many time periods. However, the forecasts of the time series econometrician also do well at many times. The professional forecasts and the no change forecasts are less commonly chosen by DMS, but even they become predominant occasionally. However, our methods allow us to see some interesting time variation in forecast performance. A general pattern that emerges (with some exceptions) is that consumer surveys forecast best in stable times. The times when the consumers are forecasting poorly are the early 1980s, 1990, 2005 and post-2007. With the exception of 2005, these are all recessionary or volatile times.

In the periods when the consumer survey forecasts receive low probability in Figure 2, it is typically the time series econometrician's forecasts which are chosen. The one exception to this pattern is the disinflationary period of 2008-2009 associated with the financial crisis and subsequent recession. This is the one period where the professional forecasts and no change forecasts (receiving roughly equal probability) are being selected by the switching state space model. An examination of the original data (see Figure 1), reveals that both of these forecasts adjusted more rapidly to the disinflation which occurred at this time. However, by the middle of 2009 the professionals and the naive forecasts are again being beaten by the consumer survey and the time series econometrician.
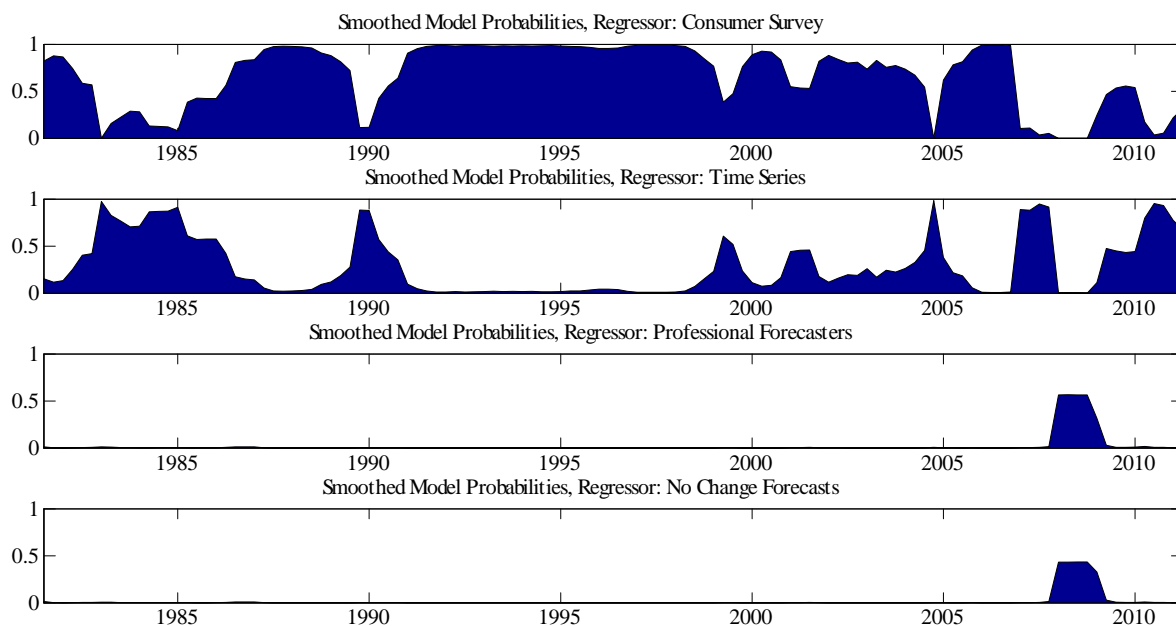
Figure 2

The model probabilities in Figure 2 are smoothed estimates based on the entire sample. Such a figure is of interest for a retrospective analysis where the researcher looks back on past forecast performance. It is also interesting to present filtered estimates, $\Pr\left(s_t = k|y^t\right)$, so as to show which forecasts a researcher in time t (given information available at time t) would have thought were good ones.[3] Figure 3 presents these filtered probabilities. It can be seen that the main patterns in Figure 3 are broadly similar to Figure 2. The consumer survey tends to receive the most probability, followed by the forecasts of the time series econometrician. Periods where the consumer survey is not selected are usually recessionary or unstable times. The main difference is post-2007 where the filtered probabilities are very erratic and indicate the no change forecasts would never have been chosen. The professional forecasters have a brief period of higher probability shortly after the financial crisis, but it is shorter than in Figure 2.

---

[3]The filtered probabilities are calculated by repeatedly running the MCMC algorithm on an expanding window of data.
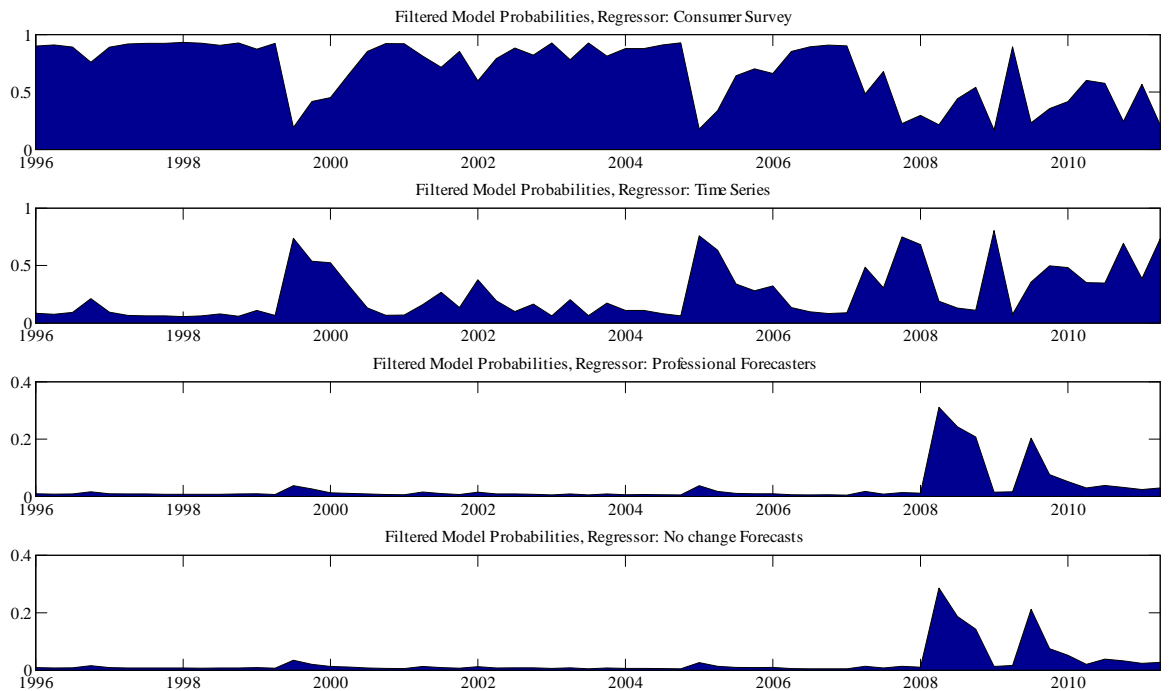
Figure 3: Using Switching State Space Model

For the sake of brevity, we do not present parameter estimates but note only that they are reasonable (e.g. the elements of θ_t tend to be around one, although there is some fluctuation over time). Having established that the switching linear state space model does seem to be a sensible way of doing model selection or model averaging in a time varying manner, we turn to the question of how it compares to the popular DMA algorithm of Raftery et al (2010).

3.3.1 Comparison to DMA and DMS using Forgetting Factors

Raftery et al (2010) introduce an algorithm for doing DMA or DMS which involves the use of forgetting factors. There has been a recent surge in popularity in using forgetting factor methods for model averaging with empirical researchers (see, e.g., Dangl and Halling, 2012, Koop and Korobilis, 2012, Koop and Tole, 2012, McCormick, Raftery, Madigan and Burd, 2012, Nicoletti and Passaro, 2012 and Koop and Korobilis, 2013). Raftery et al (2010, page 53) stress that their methods are not a special case of a switching state space model such as the one used in the present paper, but are closely related. The switching state space model approach involves specification of a matrix of Markov transition probabilities,

11

$\zeta_{ij} = \Pr(s_t = i|s_{t-1} = j)$. The forgetting factor approach does not do so. In cases where the number of models is very large, parsimony considerations mean fully specifying such a matrix is not sensible. The switching state space model involves use of MCMC methods. The forgetting factor approach leads to a filtering algorithm which does not use MCMC. In large model spaces, the computational burden of MCMC methods mean they need to be avoided. Nevertheless, the goal of our switching state space model and Raftery et al (2010)'s approach is the same: to obtain a method for model selection or model averaging done in a time varying manner.

The references in the preceding paragraph all contain empirical applications with large model spaces where forgetting factor methods are used. But they do not contain comparisons of the forgetting factor approach with a formal Bayesian model which allows for dynamic model change. With large model spaces this would be computationally infeasible. But with a small model space such as the one used in our empirical application such a comparison can be done. Our aim is to shed light on whether forgetting factor methods lead to similar empirical results as a formal Bayesian approach.

Since forgetting factor methods are established in the literature, we will not provide a description of them here. The reader unfamiliar with DMA using forgetting factors is referred to Raftery et al (2010) or they can read the brief description provided in Appendix B of this paper. Specification details, such as forgetting factor choices, are discussed in this appendix.

It is important to note that the algorithm of Raftery et al (2010) is a filtering algorithm. As described in Appendix B, It provides us with $\pi_{t|t,j}$ which is the probability that model j generated at time t, given information through time t. This a similar concept to $\Pr(s_t = k|y^t)$ which can be obtained from our switching Gaussian state space model by running our MCMC algorithm using data through time t. Both are model probabilities conditional on information available at time t. These are plotted in Figure 4 for the same four models used in the preceding sub-section.

Figure 4 should be compared to Figure 3. In a very broad sense, Figures 3 and 4 tell the same story. Most of the time the consumer surveys are providing the best forecasts for inflation. The times when the consumer surveys are not forecasting best tend to be unstable or recessionary times. The forecasts of the time series econometrician also receive appreciable weight, especially at the beginning and end of the sample.

However, there are many details in which Figures 3 and 4 differ. Most prominently, the professional forecasters do better in Figure 4 than they did in Figure 3. There are times, most particularly at the beginning and end of the sample, where the forgetting factor approach allocates considerable weight to their forecasts. The second interesting

difference is that the switching state space model seems more capable of capturing abrupt switches in model probabilities. For instance, there are several times in Figure 3 when the probability attached to the consumer surveys switched abruptly from near one to near zero or vice versa (e.g. around 1990, 2000 and 2005). These abrupt switches do not appear using the forgetting factor approach. However, the abrupt switch associated with the financial crisis does appear in both Figures 3 and 4.

What should the researcher using the forgetting factor approach to DMA make of our results? Insofar as the forgetting factor approach is viewed as a method intended to approximate a switching Gaussian state space model, it suggests the approximation is not bad. However, it is far from providing a close reproduction of exact results provided by a formal Bayesian model. Of course, as we stress above, DMA with forgetting factors is not simply a special case of a switching linear Gaussian state space model and cannot simply be interpreted as an approximation (e.g. where parameters are replaced by estimates). Nevertheless, since the two methods do have the same goal, it is not unreasonable to compare them and to hope that they will tell the same story. In this vein, we find our results (from a small model space) moderately encouraging for the user of DMA with forgetting factors working with a large model space.
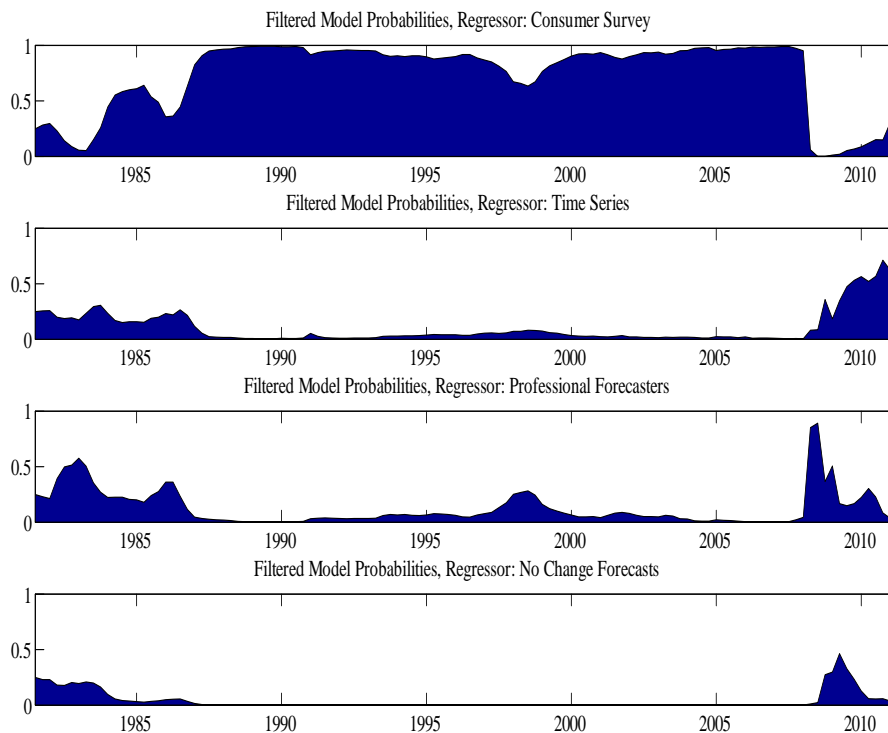
Figure 4: Using Raftery et al (2010)

3.4 Forecasting Comparison of Different Implementations of DMA/DMS

In this sub-section we compare the one-step ahead forecasting performance of our different implementations of DMA and DMS. For the switching Gaussian state space model, we repeatedly run our MCMC algorithm on an expanding window of data to provide forecasts of y_t using information available at time $t - 1$ and use this to calculate the one-step ahead predictive density. The output from this procedure can be thought of as a DMA procedure where we are averaging over $s_t = \{1, .., 4\}$. We can also obtain the predictive density for the value of s_t with highest posterior probability, a strategy analogous to DMS.

The forgetting factor approach can be used to do either DMA or DMS (see Appendix B) and involves calculating $\pi_{t|t-1,j}$ for $j = 1, .., K$ and using these probabilities either to select a single model at each point in time or average across forecasts of all models. The main results in this paper make the same forgetting factor choices as in the preceding sub-section of $\alpha = \lambda = 0.99$. However, we also present results for $\alpha = \lambda = 1$ and $\alpha = \lambda = 0.95$. The first of these Raftery et al (2010) show to be equivalent to BMA run on an expanding

window of data for a constant coefficient model. The second of these allows for a large degree of model switching and time variation in parameters. With regards to the error variance, we use both heteroskedastic and homoskedatic estimates (see Appendix B). The former of these are obtained using a rolling window of 10 observations, the latter an expanding window of data.

To evaluate our forecasts we use mean squared forecast errors (MSFEs) and mean absolute forecast errors (MAFEs) which evaluate the quality of the point forecasts. MSFEs for any approach are presented relative to the MSFE of the no change forecast. To evaluate the quality of the entire predictive density, we also present sums of log predictive likelihoods (logPL). We evaluate our forecasts over the period 1996Q1 to the end of the sample.

Several patterns emerge from Table 1. The sum of log predictive likelihoods would be the preferred Bayesian method of forecast comparison and it indicates that the forgetting factor approach forecasts slightly better than the switching linear Gaussian state space model, provided we allow for heteroskedasticity and do not choose forgetting factor values which allow for too much variation in the coefficients or too much model switching. That is, the forgetting factor method using the benchmark $\alpha = \lambda = 0.99$ choices used by Raftery et al (2010) lead to the best forecast performance (although the $\alpha = \lambda = 1$ which leads to conventional BMA in a constant coefficient model on an expanding window of data forecasts only slightly worse). Homoskedastic variants of DMA or DMS forecast quite poorly, emphasizing the importance of allowing for heteroskedasticity.

However, if we look at MSFEs and MAFEs, then a different pattern emerges where the switching linear Gaussian state space model forecasts appreciably better than forgetting factor approaches. Thus, the former methods are better at producing point forecasts. However, forgetting factor methods are clearly doing well in getting higher order moments and the entire shape of the predictive density correct.

With regards to issue of whether model averaging or model selection is better, for the switching linear Gaussian state space model, DMA methods are slightly to be preferred. However, with factor approaches, the forecast performance of DMA and DMS is very similar to one another (and, indeed, is often identical to three decimal places).

Table 1: Comparison of Forecasting Performance			
Approach	MSFE	MAFE	logPL
Switching Linear Gaussian State Space Model			
DMA	0.401	0.514	-91.48
DMS	0.431	0.551	-91.81
Forgetting Factor Approaches (heteroskedastic)			
DMA, $\alpha = \lambda = 1$	0.600	0.656	-90.89
DMA, $\alpha = \lambda = 0.99$	0.598	0.655	-90.77
DMA, $\alpha = \lambda = 0.95$	0.637	0.680	-93.10
DMS, $\alpha = \lambda = 1$	0.600	0.657	-90.89
DMS, $\alpha = \lambda = 0.99$	0.598	0.655	-90.75
DMS, $\alpha = \lambda = 0.95$	0.653	0.682	-92.46
Forgetting Factor Approaches (homoskedastic)			
DMA, $\alpha = \lambda = 1$	0.637	0.674	-110.60
DMA, $\alpha = \lambda = 0.99$	0.647	0.673	-109.53
DMA, $\alpha = \lambda = 0.95$	0.657	0.698	-105.63
DMS, $\alpha = \lambda = 1$	0.637	0.674	-110.60
DMS, $\alpha = \lambda = 0.99$	0.619	0.664	-109.52
DMS, $\alpha = \lambda = 0.95$	0.682	0.717	-109.73

In summary, our forecasting results are somewhat mixed. For the researcher interested in point forecasts, the fully Bayesian estimation procedure for switching state space models is to be preferred since it is leading to substantially lower MSFEs and MAFEs. However, for the researcher interested in the entire predictive density, DMA and DMS methods using forgetting factors are forecasting very well indicating that the approximations and compromises inherent in forgetting factor approaches do not carry a large cost with them.

3.5 Forecasting Comparison of Different Implementations of DMA/DMS in a Larger Model Space

In the preceding sub-sections, we used a small model space containing four models. However, the researcher may often wish to do DMA or DMS in larger model spaces. Accordingly, in this section, we investigate the performance of our two ways of doing DMA/DMS in the larger model space containing 15 models. That is, we augment our preceding model space with all TVP regression models containing 2, 3 or 4 explanatory variables. For the sake of brevity, we only present forecast results. However, we note that smoothed and

filtered estimates of model probabilities, comparable to Figures 2 and 3 indicate that consumer surveys and time series econometrics forecasts still dominate (sometimes individually, sometimes in the model containing the two explanatory variables π_t^{CS} and π_t^{TS}).

Table 2 presents sums of log predictive likelihoods, MSFEs and MAFEs for the same set of approaches as in Table 1. The message of Table 2 is clear. DMA or DMS done using forgetting factor methods is yielding virtually the same results in Tables 1 and 2. However, the forecasting performance of approaches based on the switching linear Gaussian state space model have deteriorated substantially.

In this empirical application, where we usually find models with a single explanatory variable to forecast best, the parsimonious forgetting factor approach is successfully choosing these models and ignoring the less parsimonious models with several explanatory variables. However, the switching linear Gaussian state space approach is not. Remember that the latter approach involves estimating a 15×15 matrix of transition probabilities. With our relatively short data set, the switching Gaussian state space model is over-parameterized and this is leading to poor forecasts.

It is worth noting, too, that the computational time associated with forecasting with this larger version of a switching Gaussian state space model is quite substantial (e.g. several hours), whereas the forgetting factor approach has a trivial computational burden (e.g. several seconds).

Table 2: Comparison of Forecasting Performance in Larger Model Space			
Approach	MSFE	MAFE	logPL
Switching Linear Gaussian State Space Model			
DMA	0.614	0.982	-113.53
DMS	0.613	0.827	-98.64
Forgetting Factor Approaches (heteroskedastic)			
DMA, $\alpha = \lambda = 1$	0.600	0.656	-90.89
DMA, $\alpha = \lambda = 0.99$	0.598	0.655	-90.77
DMA, $\alpha = \lambda = 0.95$	0.637	0.680	-93.10
DMS, $\alpha = \lambda = 1$	0.600	0.656	-90.88
DMS, $\alpha = \lambda = 0.99$	0.598	0.655	-90.75
DMS, $\alpha = \lambda = 0.95$	0.652	0.682	-92.46
Forgetting Factor Approaches (homoskedastic)			
DMA, $\alpha = \lambda = 1$	0.637	0.674	-110.60
DMA, $\alpha = \lambda = 0.99$	0.647	0.673	-109.53
DMA, $\alpha = \lambda = 0.95$	0.657	0.698	-105.63
DMS, $\alpha = \lambda = 1$	0.637	0.674	-110.60
DMS, $\alpha = \lambda = 0.99$	0.619	0.664	-109.52
DMS, $\alpha = \lambda = 0.95$	0.682	0.717	-109.73

4 Conclusions

The Bayesian empirical researcher often faces a trade-off between the desire to work with a fully specified Bayesian model and the computational burden that use of MCMC methods imposes. In the DMA literature, when the researcher works with large model spaces, it is common to use forgetting factor methods because the computational burden of doing MCMC is simply too great. In this paper, we have worked with relatively small model spaces (where MCMC methods are computationally feasible) to investigate the possible consequences of using approximate forgetting factor methods. We set up a fully specified Bayesian approach, using a switching linear Gaussian state model, which allows for model switching or model averaging in time-varying parameter models. This can be thought of as an alternative to doing DMA using forgetting factor methods.

In a small empirical application involving inflation forecasting using four models, our overall conclusion is that forgetting factor methods and the switching Gaussian state space model are leading to similar empirical results. In terms of model selection, both

approaches indicate that the consumer survey provides the best forecasts of inflation most of the time. However, both of them find that time series forecasts and surveys of professionals do tend to forecast better at particular periods (e.g. the recent financial crisis and subsequent recession). In terms of forecasting, the two approaches exhibit a similar performance if we use sums of log predictive likelihoods as a metric. However, MSFEs and MAFEs show a deterioration in forecasts for forgetting factor approaches.

When we move to a larger model space of 15 models, the forecasting performance of the switching linear Gaussian state space model deteriorates substantially. This contrasts with the forgetting factor approach where forecast performance is unaffected by the move to a larger model space. Thus, the switching Gaussian state space model can become over-parameterized even with model spaces of this size.

There are many applications [e.g. Koop and Korobilis (2012, 2013)] where the model space is so large that the use of MCMC methods is impossible. The only feasible way of doing DMA or DMS would involve some sort of approximation such as that involved in the forgetting factor approach. Insofar as the results of the present paper (based on relatively small model spaces), extend to large model spaces, they should provide reassurance that forgetting factor methods are providing reasonable results.

References

Ang, A., Bekaert, G. and Wei, M. (2007). "Do macro variables, asset markets, or surveys forecast inflation better?" Journal of Monetary Economics 54, 1163-1212.

Bates, J. and Granger, C. (1969). "Combination of forecasts," Operational Research Quarterly, 20, 451-468.

Billio, M., Casarin, R., Ravazzolo, F. and van Dijk, H. (2011). "Combining predictive densities using nonlinear filtering with applications to US economics data," Tinbergen Institute Discussion Paper, TI 2011, 172/4.

Chan, J. and Jeliazkov, I. (2009). "Efficient simulation and integrated likelihood estimation in state space models," International Journal of Mathematical Modelling and Numerical Optimisation 1, 101-120.

Chan, J., Koop, G., Leon-Gonzalez, R. and Strachan, R. (2012). "Time varying dimension models," Journal of Business and Economic Statistics, 30, 358-367.

Chipman, H., George. E. and McCulloch, R. (2001). "The practical implementation of Bayesian model selection," IMS Lecture Notes Monograph Series, 38, 66-134.

Cogley, T. and Sargent, T. (2005). "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.," Review of Economic Dynamics 8, 262-302.

Cogley, T., Morozov, S. and Sargent, T. (2005). "Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control* 29, 1893-1925.

D'Agostino, A., Gambetti, L. and Giannone, D. (2011). "Macroeconomic forecasting and structural change," Journal of Applied Econometrics, DOI: 10.1002/jae.1257.

Dangl, T. and Halling, M. (2012). "Predictive regressions with time varying coefficients," *Journal of Financial Economics,* forthcoming.

Faust, J. and Wright, J. (2012). "Forecasting inflation," chapter in forthcoming Handbook of Forecasting.

Fruhwirth-Schnatter, S. (2001a). "Fully Bayesian analysis of switching Gaussian state space models," Annals of the Institute of Statistical Mathematics 53, 31-49.

Fruhwirth-Schnatter, S. (2001b). "Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models," Journal of the American Statistical Association 96, 194-209.

Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models* (New York: Springer).

Guidolin, M. and Timmermann, A. (2009). "Forecasts of US short-term interest rates: A flexible forecast combination approach," Journal of Econometrics 150, 297-311.

Granger, C. (2006). "Invited review: Combining forecasts - twenty years later," Jour-

nal of Forecasting, 8, 167-173.

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). "Bayesian model averaging: A tutorial," Statistical Science 14, 382-417.

Kim, C.J. (1994). "Dynamic linear models with Markov switching," Journal of Econometrics 60, 1-22.

Kim, C.J. and Nelson, C. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications.* Cambridge, MA: MIT Press.

Koop, G. and Korobilis, D. (2012). "Forecasting inflation using dynamic model averaging," International Economic Review, 53, 867-886 .

Koop, G. and Korobilis, D. (2013). "Large Time-varying Parameter VARs," Journal of Econometrics, forthcoming.

Koop, G., Leon-Gonzalez, R., Strachan, R. (2009). "On the evolution of the monetary policy transmission mechanism," Journal of Economic Dynamics and Control 33, 997-1017.

Koop, G. and Tole, L. (2012). "Forecasting the European carbon market," Journal of the Royal Statistical Society, Series A, DOI: 10.1111/j.1467-985X.2012.01060.x.

Korobilis, D. (2012). "Assessing the transmission of monetary policy shocks using dynamic factor models," Oxford Bulletin of Economics and Statistics, DOI:10.1111/j.1468-0084.2011.00687.

McCormick, T., Raftery, A., Madigan, D. and Burd, R. (2012). "Dynamic logistic regression and dynamic model averaging for binary classification," Biometrics, 68, 23-30.

Nicoletti, G. and Passaro, R. (2012). "Sometimes it helps: The evolving predictive power of spreads on GDP," European Central Bank Working Paper, no. 1447.

Primiceri. G. (2005). "Time varying structural vector autoregressions and monetary policy," Review of Economic Studies, 72, 821-852.

Raftery, A., Karny, M. and Ettler, P. (2010). "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill," Technometrics, 52, 52-66.

Stock, J. and Watson, M. (2010). "Modeling inflation after the crisis," Macroeconomic Policy: Post-Crisis and Risks Ahead, FRB Kansas City symposium, Jackson Hole, Wyoming, August 26-28, 2010.

Appendix A: Bayesian Inference in the Switching Linear Gaussian State Space Model

The switching linear Gaussian state-space model the we adopt is of the form:

$$p\left(s_1 = k\right) = \tfrac{1}{K}$$
$$\xi_{jk} = p\left(s_t = k | s_{t-1} = j\right)$$
$$\theta_1 \sim N\left(0_K, 2I_K\right)$$
$$\theta_t = \theta_{t-1} + \eta_t$$
$$y_t = Z_t G^{[s_t=k]}\theta_t + \varepsilon_t,$$

for $t = 1, .., T$ and $j, k = 1, .., K$. Error assumptions and definitions of $s_t \in \{1, .., K\}$, y_t, Z_t and $G^{[s_t=k]}$ are given in Section 2. The remaining parameters of the model are $\psi = (\sigma_{\eta 1}^2, \ldots, \sigma_{\eta K}^2, \sigma_\varepsilon^{2[1]}, \ldots, \sigma_\varepsilon^{2[K]}, \xi_{11}, .., \xi_{KK})'$. We adopt a notational convention for data and states such that subscripts denote a particular time period and superscripts denote all periods up to that time period. For instance, $s^t = (s_1, .., s_t)'$ denotes all regime indicators up to time t.

We use the Gibbs sampler that sequentially draws from $p\left(\theta^T | y^T, s^T, \psi\right)$, $p\left(s^T | y^T, \theta^T, \psi\right)$ and $p\left(\psi | y^T, s^T, \theta^T\right)$. This technical appendix briefly describes each of these conditional posterior densities. The time-varying parameters are drawn from $p\left(\theta^T | y^T, s^T, \psi\right)$ using the algorithm of Chan and Jeliakov (2009). And $p\left(s^T | y^T, \theta^T, \psi\right)$ is drawn as in Fruhwirth-Schnatter (2001a,b). We refer the reader to page 420 of Fruhwirth-Schnatter (2006) for specific details of implementation. Note that this algorithm delivers $p\left(y_t | s^t, \theta^t, \psi\right)$ which, when averaged over Gibbs draws, provides us with an estimate of the predictive likelihood.

For $p\left(\psi | y^T, s^T, \theta^T\right)$ we use conditionally conjugate priors which lead to the following conditional posteriors. Given inverted Gamma priors for $\sigma_{\eta k}^2$ (for $k = 1, .., K$) with prior hyperparameters c_{0k} and C_{0k} we obtain and inverted Gamma posterior with arguments:

$$c_k(S) = c_{0k} + \tfrac{T}{2}, \quad C_k(S) = C_{0k} + \tfrac{\sum_{t=1}^T \left(\theta_{k,t+1} - \theta_{k,t}\right)^2}{2}.$$

We set the prior hyperparameters to $c_{0k} = 5$ and $C_0 = [0.08, 0.148, 0.45, 0.53]$.

For $\sigma_\varepsilon^{2[k]}$ we also use inverted Gamma priors leading to inverted Gamma conditional posteriors. We set prior hyperparameters $c_{0\varepsilon}^{[k]}$ and $C_{0\varepsilon}^{[k]}$ TO $c_{0\varepsilon}^{[k]} = 5$ and $C_{0\varepsilon}^{[k]} = [0.168, 1.480, 7.2, 10.0]$, for $k = 1, .., K$. The resulting posterior has arguments

$$c_\varepsilon^{[k]}(S) = c_{0\varepsilon}^{[k]} + \tfrac{N_{kk}}{2}, \quad C_\varepsilon^{[k]}(S) = C_{0\varepsilon}^{[k]} + \tfrac{1}{2}\sum_{t:s_t=k}^T \left(y_t - Z_t G^{[s_t=k]}\theta_t\right)^2,$$

where N_{jk} counts the number of transitions from j to k. If $j = k$ it counts the number of periods spent in regime k.

Finally, let ξ be the matrix of Markov transition probabilities ξ_{jk} and let ξ_j be the jth row of this matrix. The conditional conjugate prior for each row is Dirichlet:

$$\xi_j \sim \mathcal{D}(e_{j1}, \ldots, e_{jK}), \ \ j = 1, \ldots, K.$$

Following Fruhwirth-Schnatter (2001b), we adopt a prior reflecting a belief that the probability of staying in a regime is greater than the probability of transition to a new regime. Thus, we set $e_{jj}, \ \ j = 1, \ldots, K$ corresponding to the main diagonal to be 4. Hyperparameters off the main diagonal, $e_{ji}, \ \ j \neq i$, are set to 0.33. With this choice of prior, the conditional posterior is also Dirichlet with

$$\mathcal{D}(e_{j1} + N_{j1}, \ldots, e_{jK} + N_{jK}), \ \ j = 1, \ldots, K.$$

Appendix B: Dynamic Model Averaging Using Forgetting Factors

This appendix briefly outlines the main features of the DMA algorithm of Raftery et al (2010) as we implement it in this paper.

Suppose we have $j = 1, .., K$ TVP regression models,

$$
\begin{aligned}
y_t &= Z_t^{(j)} \theta_t^{(j)} + \varepsilon_t^{(j)} \\
\theta_{t+1}^{(j)} &= \theta_t^{(j)} + \eta_t^{(j)},
\end{aligned}
$$

$\varepsilon_t^{(k)}$ is $N\left(0, H_t^{(j)}\right)$ and $\eta_t^{(k)}$ is $N\left(0, Q_t^{(j)}\right)$. We replace $H_t^{(j)}$ by a simple estimate (the sum of squared errors divided by sample size).

If $Q_t^{(j)}$ were known, then an individual model could be estimated in a straightforward manner using the Kalman filter. $Q_t^{(j)}$ appears in the Kalman filtering prediction equation. In particular, if $y^t = (y_1, .., y_t)'$, then:

$$
\theta_t^{(j)} | y^{t-1} \sim N\left(\beta_{t|t-1}, V_{t|t-1}\right),
$$

where

$$
V_{t|t-1} = V_{t-1|t-1} + Q_t^{(j)}.
$$

This is the only place where Q_t enters the Kalman filtering formulae. If the equation for $V_{t|t-1}$ is replace by:

$$
V_{t|t-1} = \frac{1}{\lambda} V_{t-1|t-1},
$$

then MCMC methods can be avoided. λ is a forgetting factor. Forgetting factors have long been used in the state space literature to simplify estimation. There are many ways of justifying the use of forgetting factors as leading to sensible approximations. For instance, their use in this context implies that observations j periods in the past have weight λ^j. An alternative way of interpreting λ is to note that it implies an effective window size of $\frac{1}{1-\lambda}$.

The contribution of Raftery et al (2010) was to develop a filtering algorithm which also used a forgetting factor and allowed for DMA or DMS to be done. These involve calculating $\pi_{t|t-1,j}$ which is the probability that model j should be used for forecasting at time t, given information through time $t-1$. DMA arises if we average forecasts over all models using $\pi_{t|t-1,j}$ as weights. DMS arises if we choose the model with the highest value

for $\pi_{t|t-1,j}$. A recursive algorithm involving $\pi_{t|t,j}$ and $\pi_{t|t-1,j}$ can be run, beginning with $\pi_{0|0,j}$ in order to provide the necessary model probabilities. Note that $\pi_{t|t-1,j}$ is a similar concept to the $\Pr(s_t = k|y^{t-1})$ used in the switching state space model and can be used in the same manner.

Raftery et al (2010) derive a model updating equation of:

$$\pi_{t|t,j} = \frac{\pi_{t|t-1,j} p_j\left(y_t|y^{t-1}\right)}{\sum_{l=1}^J \pi_{t|t-1,l} p_l\left(y_t|y^{t-1}\right)},$$

where $p_j\left(y_t|y^{t-1}\right)$ is the predictive likelihood for model j produced by the Kalman filter. However, instead of using a Markov transition matrix to model the probability of switching between models, a model prediction equation involving a forgetting factor α is used:

$$\pi_{t|t-1,j} = \frac{\pi_{t-1|t-1,j}^\alpha}{\sum_{l=1}^J \pi_{t-1|t-1,l}^\alpha}.$$

This algorithm has a large advantage in that no MCMC is required and a complete specification of a Markov transition matrix is not required. It is computationally efficient, involving only the filtering algorithms just described. Its properties are described in more detail in Raftery et al (2010).

This algorithm requires selection of α, λ, $\pi_{0|0,j}$ and initial conditions for the time varying parameters. For the last, we use the same values as for the switching state space models. For $\pi_{0|0,j}$, we use the noninformative choice of $\pi_{0|0,j} = \frac{1}{K}$. The main results in the paper set $\alpha = \lambda = 0.99$, although (as noted in the body of the paper) in our forecast comparison exercise we experiment with different values. These forgetting factors are not directly comparable to the parameters in the switching state space models. But, loosely speaking, forgetting factors and priors in switching state space models are both chosen to allow for a moderate degree of parameter and model change.