Model Switching in Time-Varying Parameter Regression Models*

Miguel Belmonte
University of Strathclyde

Gary Koop
University of Strathclyde

August 2012

Abstract: This paper investigates the usefulness of switching Gaussian state space models as a tool for implementing dynamic model selection in time-varying parameter regression models. Dynamic model selection 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. We compare our exact approach to dynamic model selection to a popular existing procedure which relies on the use of forgetting factor approximations. In an application, we investigate which of several different forecasting procedures works best for inflation. We also investigate whether the best forecasting method changes over time.

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

Preliminary and incomplete paper: Please do not quote

*This research was supported by the ESRC under grant RES-062-23-2646. Gary Koop is a Fellow at the Rimini Centre for Economic Analysis. Address for correspondence: Gary Koop, Department of Economics, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, UK. Email: Gary.Koop@strath.ac.uk

1 Introduction

Bayesian model averaging or model selection (BMA or BMS) 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 $Z^{(k)}$ 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 $N\left(0, \Sigma_\eta^{(k)}\right)$.

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 $\Pr(s_t = k|y^{t-1})$. DMA involves averaging across models using these probabilities. Different approaches to DMA or DMS arise when different models or methods are used to calculate $\Pr(s_t = k|y^{t-1})$. Raftery et al (2010), working in an application involving many explanatory variables and, hence, a large model space, uses forgetting factor methods to approximate $\Pr(s_t = k|y^{t-1})$. 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 $\Pr(s_t = k|y^{t-1})$ for $t = 1, .., T$ as opposed to $\Pr\left(s_t = k|y^T\right)$.

The purpose of this paper is to propose 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 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 use 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.

This paper contains an application involving selecting between 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 [insert preliminary results when available].

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 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 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

$i, j = 1, .., K$. x_t is a vector of unobserved states.

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 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 k^{th}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.[1] 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 $\sigma^2_{\eta k}$ so that the regression coefficients evolve independently of one another.

In 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 k^{th} explanatory variable is used. Switches 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 this algorithm (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

[1] This definition of $G^{[s_t=k]}$ makes sense in our application. But other choices could be made in different applications without altering the basic algorithm.

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,

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

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.

[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.

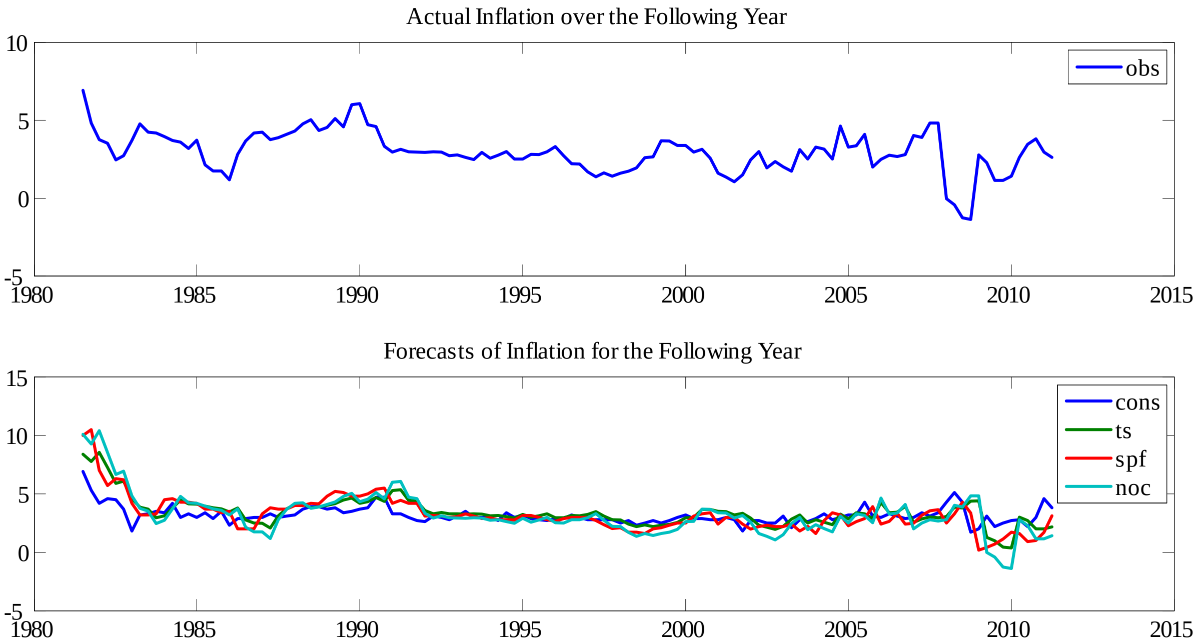

Figure 1

In terms of the notation used in Section 2, $y_t = \pi_t^R$ and $Z_t = \left(\pi_t^{SPF}, \pi_t^{CS}, \pi_t^{TS}, \pi_t^{NOC}\right)$. All variables are in deviation from mean form and we do not include an intercept.

3.3 Smoothed Estimates using the Switching Gaussian State Space Approach

Our main interest is in which of the four forecasts has been best at each point in time. To shed light on this, we begin by presenting smoothed estimated of the regime probabilities, $\Pr\left(s_t = k|y^T\right)$ for $k = 1, .., 4$ and $t = 1, .., T$.

Our results confirm the general findings of Ang, Bekaert and Wei (2007) and Faust and Wright (2012). Surveys, regardless of whether of consumers or professionals, do tend to forecast better than either the time series econometrician or the naive agent. Simple OLS regression methods also confirm this general finding. If we run a regression of y_t on each of the inflation forecasts, we find it to be $0.132, 0.104, 0.185$ and 0.099, respectively forπ_t^{CS}, π_t^{TS}, π_t^{SPF} and π_t^{NOC}, respectively. However, our methods allow us to see some interesting time variation in forecast performance.

The four panels in Figure 2 must, by definition, sum to one. It can be seen that it is rarely the case for one model to be completely dominant (i.e. have $\Pr\left(s_t = k|y^T\right)$ to be near one for any k). However, typically if the probability associated with the models using the consumer and professional surveys are added together, they are dominant. However, there is one interesting exception to this. At the time of the financial crisis, the naive no-change forecast dominates. It caught the steep fall in inflation that occurred in the recession better than any of the other approaches. The forecast of the time series econometrician is particularly bad. Our methodology never allocates appreciable probability to π_t^{TS}.

As to the question of whether professionals or consumers forecast better, it can be seen that the answer depends on the time period. Throughout much of the 1990s (a stable period) the consumer survey forecasts best. In the 1980's, though, the professionals are forecasting better. After 1996, there is a lot of switching between the two forecasts. And, in the time of the financial crisis and shortly thereafter, both of the forecasts are beaten by the no change forecast.

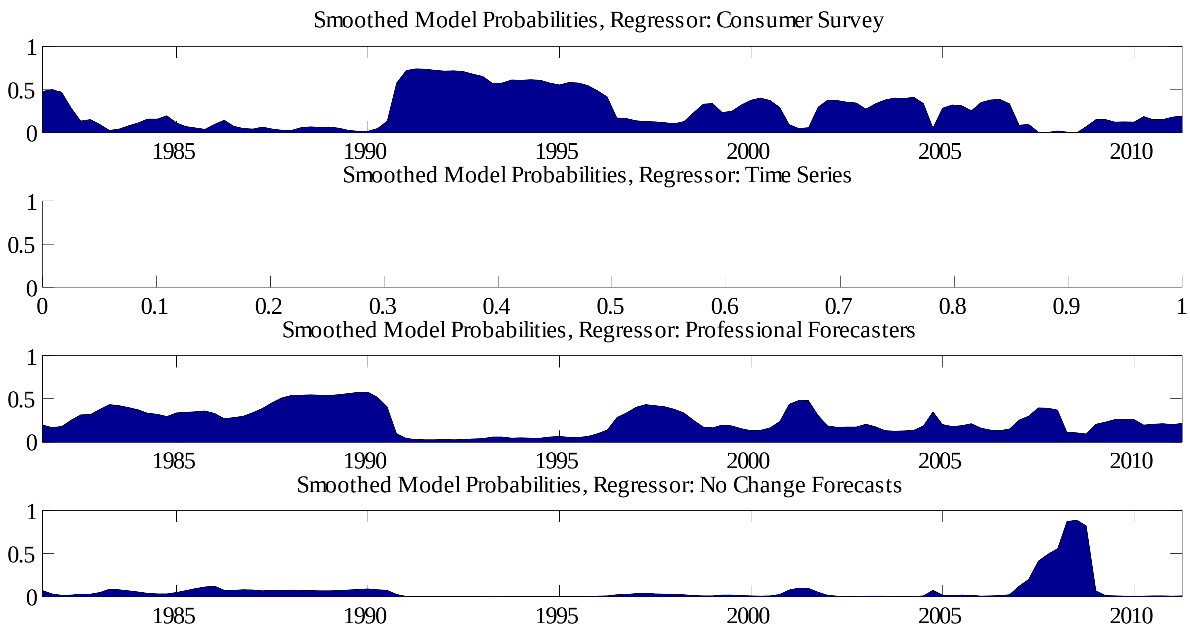

Figure 2

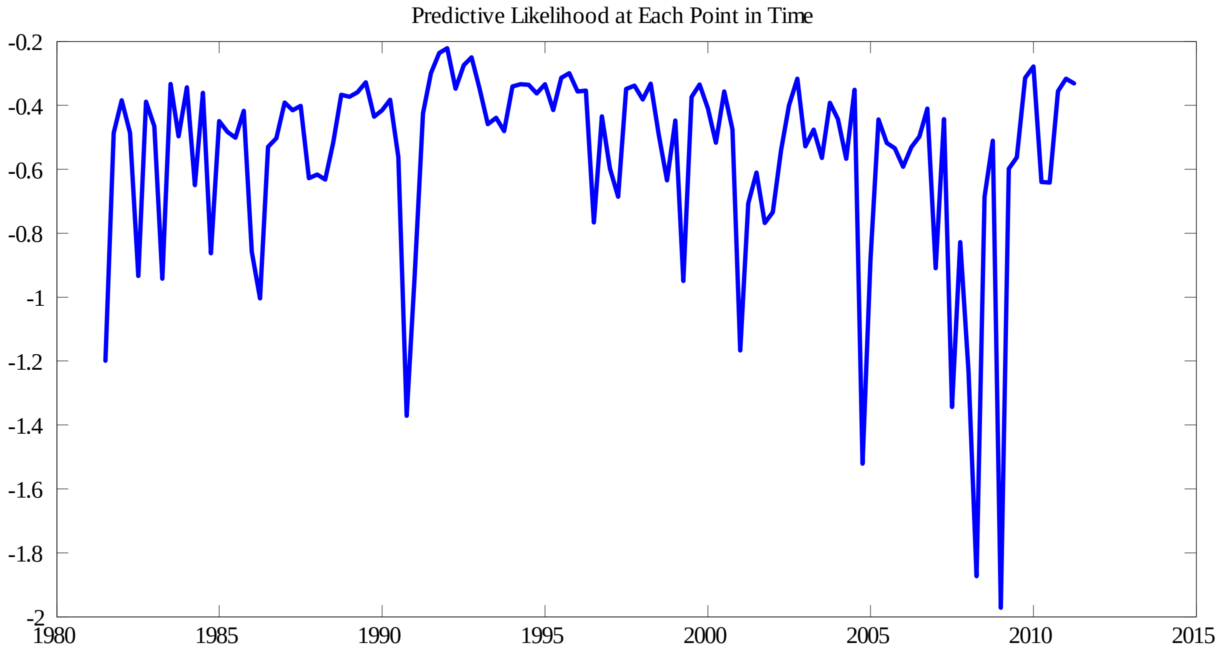

Figure 3

Figure 3 plots the predictive likelihood for the switching Gaussian state space model over time (see Technical Appendix for details). It can be seen that the performance of the model does vary substantially over time. Unsurprisingly, it performs worst in the recent financial crisis. But there are also other times of poor performance such as the early 1990s. It is interesting to note that these times of poor performance are also times where we tend to find model switching. That is, the financial crisis was the time the model switched to choosing the no-change forecast and the early 1990s was when it switched from choosing the professional to consumer surveys. This indicates that, when forecast performance is deteriorating,

More empirical results will follow.

3.3.1 Comparison to DMS using Methods of Raftery et al (2010)

This section remains to be completed.

4 Conclusions

To be completed.

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 .

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," Journal 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, 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, forthcoming.

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.

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.

Technical Appendix

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

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

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 $\psi=(\sigma^2_{\eta 1},\ldots,\sigma^2_{\eta K},\sigma^{2[1]}_{\varepsilon},\ldots,\sigma^{2[K]}_{\varepsilon},\xi_{11},..,\xi_{KK})'$. 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). Note that this algorithm provides us with an estimate of $p\left(y_t|s^t,\theta^t,\psi\right)$ 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 the following conditional posteriors. Given inverted Gamma priors for $\sigma^2_{\eta k}$ (for $k=1,..,K$) with prior hyperparameters c_{0k} and C_{0k} we obtain and inverted Gamma posterior with arguments:

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

For $\sigma^{2[k]}_{\varepsilon}$ we also use inverted Gamma priors leading to inverted Gamma conditional posteriors. Given prior hyperparameters of $c^{[k]}_{0\varepsilon}$ and $C^{[k]}_{0\varepsilon}$ for $k=1,..,K$, the posterior has arguments

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

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 j^{th} row of this matrix. The prior for each row is assumed to be Dirichlet:

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

With the prior, the conditional posterior is also Dirichlet with

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