Bayesian Analysis (2011) **6**, Number 4, pp. 533–540

# Comment on Article by Wyse et al.

Gary Koop*

## 1 Introduction

Computation can be a serious challenge in multiple changepoint models especially when simulation methods are used. This paper, building on earlier contributions such as Fearnhead (2006), makes an important contribution in addressing this challenge. The main contribution of the present paper is, through the use of INLAs, to extend the feasibility of such methods to a wider class of likelihood functions by allowing for data that is dependent within segments. An additional contribution is the development of reduced filtering recursions which further decreases the computational burden. These methods will be found useful in a wide variety of empirical applications.

In economics there is a large changepoint (or structural break) literature. Papers such as Stock and Watson (1996), Ang and Bekaert (2002) and Bauwens et al (2011) find empirical evidence of widespread parameter instability in macroeconomic and financial time series. Clements and Hendry (1998) argue that structural breaks are the main reason for forecast failure. Pesaran, Pettenuzzo and Timmermann (2006), Koop and Potter (2007) and Maheu and Gordon (2008), among many others, develop forecasting methods for changepoint models. My comments will focus on the issue of how the methods proposed in the paper can be used in the context of this economic literature.

## 2 Forecasting with Changepoint Models

My reading of the experience of economists can be summed up with the phrase: priors are important. We require priors for $\theta_j$ (the parameters which characterize the likelihood in regime $j$), $\tau_j$ for $j = 1, .., k$ (the changepoints) and $k$ (the number of changepoints).[1] Each of these is important and my comments will deal with each in turn.

*Department of Economics, University of Strathclyde, Glasgow, UK, Gary.Koop@strath.ac.uk

[1]Throughout this note, I use the same notation as in the paper itself.

 DOI:10.1214/11-BA620B

## 2.1 Priors for Parameters in Each Regime

We often face the situation where there are many changepoints and our sample size is only moderately large. In this case, there can be few observations in each regime and priors can have a big impact. In addition, the goal of the forecaster with forecast horizon $h$ is to obtain $p\left(y_{t+h}|y_{1:t}\right)$. Let us suppose $\tau_{j-1} \leq t \leq \tau_j$. In a changepoint model involving prior independence over $\theta_j$, the predictive density reduces to $p\left(y_{t+h}|y_{\tau_{j-1}:t}\right)$. If $\tau_{j-1}$ is close to $t$, this predictive density is based on very few observations and the importance of the prior becomes particularly important. Even worse, if a changepoint occurs between $t$ and $t+h$, then (depending on the structure of the model) the researcher will either end up forecasting using the incorrect $\theta_j$ or, if the correct $\theta_{j+1}$ is used to forecast, it must be drawn from the prior.

For these reasons, many economic forecasters use hierarchical priors. The following quotation, from a financial application involving the equity premium, encapsulates the thoughts of many economists nicely.

> "In standard approaches to models that admit structural breaks, estimates of current parameters rely on data only since the most recent estimated break. Discarding the earlier data reduces the risk of contaminating an estimate of the equity premium with data generated under a different [process]. That practice seems prudent, but it contends with the reality that shorter histories typically yield less precise estimates. Suppose... a shift in the equity premium occurred a month ago. Discarding virtually all of the historical data on equity returns would certainly remove the risk of contamination by pre-break data, but it hardly seems sensible in estimating the current equity premium. Completely discarding the pre-break data is appropriate only when the premium might have shifted to such a degree that the pre-break data are no more useful ..., than, say, pre-break rainfall data, but such a view almost surely ignores economics." Pastor and Stambaugh (2001, pages 1207-1208).

The hierarchical priors used by economists take two main forms: i) $\theta_j$ are independent draws from a common distribution [e.g. Pesaran, Pettenuzzo and Timmermann (2006) and Song (2011a)'s treatment of the Maheu and Gordon (2008) approach]; ii) the prior mean for $\theta_j$ depends on $\theta_{j-1}$ [e.g. time-varying parameter state space models or Koop and Potter (2007)].

The approach outlined in the paper assumes $\theta_j$ and $\theta_s$ are, a priori, independent for $j \neq s$. Of course, in many cases (particularly those not involving forecasting and where segment durations tend to be long) having such an independent prior is sensible. But many economists would be very interested in seeing whether and how the methods of this paper could be extended to deal with hierarchical priors of the sort just described. Fearnhead and Liu (2011) treats the case where the prior for $\theta_j$ depends on $\theta_{j-1}$ and the present paper could be easily extended to handle this case. Treatment of this and other richer categories of hierarchical prior would be an interesting extension of the present approach.

## 2.2 Priors for the Changepoints

Priors over $\tau_1, .., \tau_k$ can also be important and, particularly in the context of a model with a fixed number of changepoints, apparently sensible priors can have undesirable properties. This point has been drawn out in several papers, including one of my own: Koop and Potter (2009). Let me illustrate it using a simple example (although the point I will make holds much more generally), suppose the researcher has a model with two changepoints and wishes to use an apparently "noninformative" uniform prior over them. Such a prior is sometimes (either implicitly or explicity) written as $p(\tau_1, \tau_2) = p(\tau_1)\, p(\tau_2|\tau_1)$ where both of the distributions on the right hand side are uniform:

$$p(\tau_1) = \frac{1}{n-2} \text{ for } \tau_1 = 1, ..., n-2$$

$$p(\tau_2|\tau_1) = \frac{1}{n-\tau_1-1} \text{ for } \tau_2 = \tau_1 + 1, ..., n-1.$$

It is easily shown that the marginal prior, $p(\tau_2)$, is very different from the "noninformative" uniform distribution. Prior probability for $\tau_2$ can pile up near the end of the sample (exactly at the most crucial point for the forecaster). In the present paper, the prior is written as $p(\tau_1, \tau_2) = p(\tau_1|\tau_2)\, p(\tau_2)$ and if one were to use similarly uniform priors the prior probability would occur at the beginning of the sample.

Of course, the approach outlined in the paper holds for a much wider class of priors. The preceding example is partially meant just to illustrate the importance of prior choice with respect to the changepoints. But it is leading me to what is potentially a more important issue: treatment of $k$, the number of changepoints. Before I discuss this , note that it is often more natural to think in terms of the distribution of the durations

($d_j = \tau_j - \tau_{j-1}$) of the segments, rather than in terms of the distribution of changepoints. Just as economists find it useful to use hierarchical or dependent priors for $\theta_j$, it is also potentially useful to have hierarchical or dependent priors for $d_j$. Economists often give names to their segments (e.g. expansions and recessions in the macroeconomy or bull and bear markets in equities) and such naming suggests they may have some common properties. In such cases, information about the duration of past segments can be useful in obtaining more precise inferences about the duration of current segments. For instance, when forecasting, information about the duration of past regimes can be helpful in predicting the duration of the current regime (and, thus, predicting whether a changepoint will occur at the time being forecast). Having a prior which exhibits dependence between $d_j$ and $d_s$ for $j \neq s$ (in some appropriate manner which would depend upon the application at hand) is an attractive way of modelling such properties. I have found that models with such features are often very computationally demanding using MCMC methods. I expect many economists would find a computationally efficient approach to models with such priors on the durations attractive. Unlike incorporation of hierarchical priors for $\theta_j$ (which looks like a simple extension of the present approach), allowing for prior dependence between $d_j$ for $j = 1, .., k$ looks like a more challenging extension of the simulation-free approach outlined in the paper.

## 2.3 Priors for the Number of Changepoints

The treatment of $k$ is closely connected with the prior for the durations. Many papers simply treat $k$ as fixed. Alternatively, the present paper (and many others) estimate a model conditional on $k$, but then calculate the marginal likelihood for different values of $k$ in order to build up a posterior for $k$. In many cases, this is a sensible thing to do. But it does have implications which may be undesirable. It would be interesting to see if the computationally efficient approach developed in this paper could be adapted to deal with models which do not have this property.

Note that the algorithm fixes $\tau_0 = 0$ and $\tau_{k+1} = n$, ensuring that precisely $k$ changepoints occur in-sample. One issue this raises is that this limits the kinds of prior distributions one can use. For instance, the Poisson distribution is a flexible one that has been used to model durations. But is unbounded. So if one were to use it as a prior for durations, there would be no guarantee that one changepoint would occur before time $n$, much less $k$. Approaches such as Maheu and Gordon (2008) do not impose a fixed number of breaks and allow for priors such as the Poisson for the durations.

The pathologies in my simple example in the preceding sub-section were due to the fact that the prior imposed exactly $k = 2$ changepoints would occur by time $n$. To continue this simple example, note that we can construct a prior without those pathologies (where the marginal priors for both changepoints are uniform) by replacing $p(\tau_2|\tau_1)$ specified there with

$$p(\tau_2|\tau_1) = \frac{1}{n-2} \text{ for } \tau_2 = \tau_1 + 1, ..., n + \tau_1 - 2.$$

But note that this prior does not ensure that the second changepoint occurs before time $n$. The second changepoint occurs out of sample. In Koop and Potter (2009), we argue that, for the informative prior Bayesian, posteriors and predictives will still be proper densities. And, indeed, allowing for $k$ changepoints, all of which could occur either in-sample or out-of-sample, allows the researchers to estimate the number of changepoints which occur in-sample. Hence, using priors with such properties may be attractive.

Another attractive approach to estimating the number of changepoints is developed in papers such as Song (2011b). A nonparametric approach (involving a Dirichlet process prior) is used to model the probabilities of switching between segments. This approach allows for the number and nature (e.g. whether segments are recurring or non-recurring) of segments to be estimated jointly in one statistical model. Jochmann (2010) develops a similar approach also using an infinite hidden Markov model.

The approaches outlined in this sub-section do have some empirically attractive properties. However, they can be computationally demanding since they require the use of MCMC methods. It would be interesting to see if the computationally efficient algorithm proposed in this paper could be extended to cover them.

# 3 Conclusion

In their paper, the authors have convinced me that the algorithm they propose is accurate and computationally efficient and potentially of use in changepoint problems. However, it is developed in the context of a particular set of priors. In my discussion, I have argued that this set of priors excludes some that are of interest to the economist and speculated on whether the algorithm could be adapted to handle them. In some cases, I have concluded that adaptations would be simple, in others more difficult. In all cases, though, they would be interesting.

## References

Ang, A. and Bekaert, G. (2002). "Regime switches in interest rates." *Journal of Business and Economic Statistics*, 20: 163–182.

Bauwens, L., Koop, G., Korobilis, D., and Rombouts, J. (2011). "A comparison of forecasting procedures for macroeconomic series: The contribution of structural break models." Discussion paper 3, Centre for Operations Research and Econometrics.

Clements, M. and Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press: Cambridge.

Fearnhead, P. (2006). "Exact and efficient Bayesian inference for multiple changepoint problems." *Statistics and Computing*, 16: 203–213.

Fearnhead, P. and Liu, Z. (2011). "Efficient Bayesian analysis of multiple changepoint models with dependence across segments." *Statistics and Computing*, 21: 217–229.

Jochmann, M. (2010). "Modeling U.S. inflation dynamics: A Bayesian nonparametric approach." Working paper 03_10, Rimini Centre for Economic Analysis.

Koop, G. and Potter, S. (2007). "Estimation and forecasting in models with multiple breaks." *Review of Economic Studies*, 74: 763–789.

— (2009). "Prior elicitation in multiple change-point models." *International Economic Review*, 50: 751–772.

Maheu, J. and Gordon, S. (2008). "Forecasting, learning and structural breaks." *Journal of Applied Econometrics*, 23: 553–583.

Pastor, L. and Stambaugh, R. (2001). "The equity premium and structural breaks." *Journal of Finance*, 56: 1207–1239.

Pesaran, M., Pettenuzzo, D., and Timmermann, A. (2006). "Forecasting time series subject to multiple structural breaks." *Review of Economic Studies*, 73: 1057–1084.

Song, Y. (2011a). "An efficient approach to estimate and forecast in the presence of an unknown number of change-points." Chapter 2 of PhD dissertation, University of Toronto.

— (2011b). "Modeling regime switching and structural breaks with an infinite dimension Markov switching model." Chapter 3 of PhD dissertation, University of Toronto.

Stock, J. and Watson, M. (1996). "Evidence on structural instability in macroeconomic time series relations." *Journal of Business and Economic Statistics*, 14: 11–30.