Table 3: Forecasting Comparison

	MB	Performance Relative to MB			
		MNR (MSC)	MS (MSC)	TVP	Rolling OLS(40)
In Sample					
AIC difference	–447.8	–12.2	–12.2	0.3	
Out of sample: No Re-estimation					
KL divergence	–215.3	21.9*	25.0*	23.2*	5.5
		(3.4)	(2.4)	(3.3)	(0.8)
Relative MSFE	6.60	1.16	1.62*	1.55*	1.04
		(1.73)	(3.69)	(3.62)	(0.50)
Out of sample: Recursive Estimation					
KL divergence	–214.9	22.2*	15.5*	4.1	5.8
		(2.0)	(2.0)	(1.0)	(0.8)
Relative MSFE	5.68	1.35	1.47*	1.04	1.21
		(1.8)	(2.1)	(0.6)	(1.3)

Note: I estimate the model in (**) for h=2 using the sample 1967:Q1-1986:Q4, and forecast over the period 1987:Q1-2009:Q4. The MB column shows AIC (in sample), predictive log likelihood (out of sample), and mean squared forecast error (out of sample). Defining K as the number of estimated parameters, $AIC = 2L(\theta) - 2K$. Below the out-of-sample statistics in parentheses are t-statistics for testing a zero difference between the MB model and the alternative model; a * superscript denotes significance at 5 percent using standard normal critical values (West 1996).

4.3 Forecasting Comparison

In Table 3, I present an out-of-sample forecasting comparison of the MB model to several alternative breaks models for the two-quarter horizon model. I estimate (18) using data up to 1986:Q4, and use the data from 1987–2009 to evaluate post-sample forecasting performance. I compare the MB model to Markov switching models with recurring (MS) and nonrecurring (MNR) states, a time-varying parameter GARCH(1,1) model, and 10-year rolling OLS regressions. Table 3 shows the estimated Kullback-Leibler information loss from applying these alternative models rather than an MB model. I estimate the KL loss using AIC (Akaike 1973) for the estimation sample and the predictive likelihood for the post sample forecasting period (Cooley and Parke 1990). In addition, I present in Table 3 the mean-squared forecast errors (MSFE) of the alternative models relative to the MB model for the out-of-sample period.

I conduct the forecasting experiment using both fixed and expanding estimation samples, with the exception of the MNR model and the rolling OLS regressions. The MNR model has no capacity to predict post-break values of β_t and σ_t, so I re-estimate the parameters of this model every quarter, and I select the number of states using the Markov switching criterion (MSC) of Smith, Naik and Tsai (2006). I also use MSC to select the number of states in the MS model. For all models, the forecasts should be interpreted as one-quarter-ahead predictions because, when forecasting period t+1, the filtering algorithms use information up to period t to infer the coefficient values even though the explanatory variable is measured at period t-1.

I perform the forecasting exercise with a fixed estimation sample to highlight the ability of the models to adapt to breaks, which is an ability that each model is designed to possess. Table 3 shows that the MS and TVP-GARCH models perform markedly worse in such a comparison, with predictive log likelihood values 25.0 and 23.2 points worse than the MB model. Similarly, their MSFE values are 62 percent and 55 percent worse than the MB model. Even the MNR and rolling OLS models, which require re-estimation to generate forecasts, perform worse than the fixed-sample MB model. The MNR model is 21.9 points worse in predictive log likelihood and 16 percent worse by MFSE. The rolling OLS model gets closest, with predictive log likelihood 5.5 points worse and MSFE 4 percent worse than the MB model.

Recursive re-estimation of MB model increases its performance advantage. Its RMSE decreases from 6.60 to 5.68, which is 35 percent better than the recursively estimated MNR and 47 percent better than the recursively estimated MS models. The MB model improves to 21 percent better than the rolling OLS estimates, and it is 4 percent better than the recursively estimated TVP model. Recursive estimation produces little improvement in the predictive log likelihood of the MB model, but it remains substantially better than the recursively estimated MS and MNR models. Overall, the MB model outperforms the other models out of sample, although the difference is not statistically significant at 5 percent for the rolling OLS models.

5. Conclusion

In this article I develop the MB model for estimation and forecasting in regressions with changing coefficients and error variances. I parameterize the model using a two-state hidden Markov process, which allows me to apply the standard Markov switching filter and to keep the state space of low dimension. Evaluating the likelihood in a particular period requires knowledge only of the most recent break date. It does not require knowledge of the entire sequence of break dates up to that period, which makes the model computationally straightforward. The resulting MB model outperforms competing breaks models in an application to the predictive ability of the yield curve for GDP growth.

The MB model generates conditional parameter estimates and forecasts by averaging over models that include progressively more historical data. This feature provides a link to the forecast combination literature (Timmermann 2006), in which averaging across models often improves forecasting performance. Moreover, it explains why the model can perform well even when the breaks are small and therefore difficult to identify. Further research into the links between forecast combination and the MB model will further improve forecasting and inference in the presence of breaks and model uncertainty. See Pesaran and Pick (2011) for recent work on a similar topic.

References

Akaike, H., 1973. *Information Theory and an Extension of the Maximum Likelihood Principle*. In 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki (Eds.), 267-281. Budapest: Akademia Kiado.

Andrews, D.W.K., 2001, "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69(3): 683-734.

Andrews, D.W.K. and W. Ploberger, 1994, "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62(6): 1383-1414.

Atkeson, A. and L.E. Ohanian, 2001, "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1): 2–11.

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

Chib, S., 1998, "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86(2): 221-241.

Chou, R., R.F. Engle, and A. Kane, 1992, "Measuring risk aversion from excess returns on a stock index," *Journal of Econometrics*, 52(2): 201–24.

Clark, T.E. and M.W. McCracken, 2005, "The Power of Tests of Predictive Ability in the Presence of Structural Breaks," *Journal of Econometrics*, 124(1): 1-31.

Cogley, T. and T.J. Sargent, 2005, "Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S.," *Review of Economic Dynamics,* 8(2): 262-302.

Cooley, T.F., and W.R. Parke, 1990, " Asymptotic Likelihood-Based Prediction Functions," *Econometrica*, 58(5): 1215-34.

Cooley, T.F., and E.C. Prescott, 1973, "An Adaptive Regression Model," *International Economic Review*, 14(2): 364-371.

Davies, R.B., 1977, "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64(2): 247-54.

Douc, R. E. Moulines, J. Olsson, and R. van Handel, 2011, "Consistency of the maximum likelihood estimator for general hidden Markov models," *The Annals of Statistics*, 39(1): 474–513.

Douc, R., E. Moulines and T. Rydén, 2004, " Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime," *The Annals of Statistics*, 32(5): 2254–2304.

Efron, B., 2008, "Microarrays, Empirical Bayes and the Two-Groups Model," *Statistical Science*, 23(1): 1–22.

Elliott, G. and U.K. Müller, 2006, "Optimally Testing General Breaking Processes in Linear Time Series Models," *Review of Economic Studies*, 73(4): 907-940.

Engle, R.F. and A.D. Smith, 1999, "Stochastic Permanent Breaks," *Review of Economics and Statistics*, 81(4): 553-574.

Estrella A., A.P. Rodrigues, and S. Schich, 2003, "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States," *The Review of Economics and Statistics*, 85(3): 629–644.

Fine, S., Y. Singer and N. Tishby, 1998, "The Hierarchical Hidden Markov Model: Analysis and Applications", *Machine Learning*, 32: 41–62.

Giacomini, R. and B. Rossi, 2006, "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?," *Oxford Bulletin of Economics and Statistics*, 68(S): 783-795.

Gerlach, R., C. Carter, and R. Kohn, 2000, "Efficient Bayesian Inference for Dynamic Mixture Models," *Journal of the American Statistical Association*, 95(451): 819-828.

Giordani, P., and R. Kohn, 2008, " Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models," *Journal of Business and Economic Statistics*, 26(1): 66-77.

Hamilton, J.D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2): 357-84.

Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press: Princeton.

Harrison, P.J. and C.F. Stevens, 1976, "Bayesian Forecasting," *Journal of the Royal Statistical Society, Series B*, 38(3): 205-247.

Harville, D.A., 1977, "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72(358): 320-338.

Hildreth, C. and J.P. Houck, 1968, "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63(322): 584–95.

Hinkley, D., 1979, "Predictive Likelihood," *The Annals of Statistics*, 7(4): 718-728.

Kim S., N. Shephard and S. Chib, 1998, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65(3): 361-393.

Koop, G., 2003, Bayesian Econometrics, John Wiley & Sons: West Sussex.

Koop, G., and S. Potter, 2007, "Forecasting and Estimating Multiple Change-Point Models," Review of Economic Studies, 74(3): 763-789.

Lauritzen, S.L., 1974, "Sufficiency, Prediction, and Extreme Models," *Scandinavian Journal of Statistics*, 1: 128-134.

Leroux, B.G., 1992, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes and their Applications,* 40(1): 127–143.

McConnell, M.M. and G. Perez-Quiros, 2000, "Output Fluctuations in the United States: What Has Changed since the 1980s?" *American Economic Review*, 90(5): 1464-1476.

McCulloch, R.E., and Tsay, R.S., 1993, "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88(423): 968-978.

Pesaran, M.H, D. Pettenuzzo , a nd A. Timmermann, 2006, "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73(4): 1057-1084.

Pesaran, M.H. and A. Pick, 2011, "Forecast Combination Across Estimation Windows," *Journal of Business and Economic Statistics*, in press.

Pesaran, M.H. and A. Timmermann, 2007, "Selection of Estimation Window in the Presence of Breaks,".*Journal of Econometrics*, 137(1): 134-161.

Robbins, H., 1955, "An Empirical Bayes Approach to Statistics," *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability*, 1: 157-164.

Self, S.G. and K.Y. Liang, 1987, " Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under non-Standard Conditions," *Journal of the American Statistical Association*, 82(398): 605–610.

Smith, A., 2005, "Level Shifts and the Illusion of Long Memory in Economic Time Series," *Journal of Business and Economic Statistics*, 23(3): 321-335.

Smith, A., P.A. Naik, and C.L. Tsai, 2006, "Markov-Switching Model Selection using Kullback-Leibler Divergence," *Journal of Econometrics*, 134(2):553-577.

Stock, J.H. and M.W. Watson, 1989, "New Indexes of Coincident and Leading Indicators," *NBER Macroeconomic Annual*, 4: 351-394.

Stock, J.H., and M.W., Watson, 1996, "Evidence on S tructural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14(1): 11-30.

Timmermann, A., 2001, "Structural Breaks, Incomplete Information, and Stock Prices," *Journal of Business and Economic Statistics*, 19(3): 299-314.

Timmermann, A., 2006. "Forecast Combinations," in *Handbook of Economic Forecasting* (Edited by Elliott, G., C.W.J. Granger, and A. Timmermann) North Holland.

Vuong, Q.H., 1989, " Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57(2): 307-33.

West, K.D., 1996, "Asymptotic Inference about Predictive Ability," *Econometrica*, 64(5): 1067-84.

Wooldridge, J.M., 2005, "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models," *Review of Economics and Statistics*, 87(2): 385-390.