Formal and Heuristic Model Averaging Methods for Predicting the US Unemployment Rate*

Jeremy Kolly[1]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1]Department of Management, University of Fribourg, Switzerland
jeremy.kolly@unifr.ch

Abstract

We consider a logistic transform of the monthly US unemployment rate. For this time series, a pseudo out-of-sample forecasting competition is held between linear and nonlinear models and averages of these models. To combine predictive densities, we use two complementary methods: Bayesian model averaging and optimal pooling. We select the individual models combined by these methods with the evolution of Bayes factors over time. Model estimation is carried out using Markov chain Monte Carlo algorithms and predictive densities are evaluated using statistical tests, log scores and probability integral transforms. The sophisticated averages of linear and nonlinear models turn out to be valuable tools for predicting the US unemployment rate in the short-term.

Keywords: nonlinearity; model combination; Markov chain Monte Carlo methods; Bayes factors; forecast evaluation

1 Introduction

Many studies point out that nonlinear models are able to yield superior predictions of the US unemployment rate [10, 9, 6, 1, 12, 2]. Among them, [12] and [2] argue in favor of the logistic smooth transition autoregression (LSTAR). This nonlinear regime-switching model, proposed by [11], can be written as:

$$y_t = \phi_{10} + \sum_{j=1}^{p} \phi_{1j} y_{t-j} + G(s_t; \gamma, c)\left(\phi_{20} + \sum_{j=1}^{p} \phi_{2j} y_{t-j}\right) + \epsilon_t$$

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp[-\gamma^2 (s_t - c)]}$$

where the ϵ_t are i.i.d. $N(0, \sigma^2)$ and where the logistic transition function $G(\bullet)$ depends on the observable transition variable s_t and contains γ and c; the smoothness and location parameters respectively.

*This short paper is based on the PhD thesis of the author, comments are welcome.

In the contributions mentioned previously, the linear models are found to be good competitors. This may mean that linear and nonlinear models provide complementary descriptions of the US unemployment process. The present research takes this possibility into account by investigating the predictive performance of averages of linear and nonlinear models. Some of the above-mentioned studies consider model combination. However, their appoaches are either limited or different.

2 Model Averaging Methods

Consider the model space $\mathcal{M} = \{M_1, \dots, M_K\}$ where each model delivers a predictive density $p(y_{T+1}|y_{1:T}, M_k)$ for the future observation y_{T+1} given the sample $y_{1:T} = (y_1, \dots, y_T)'$. These predictive densities can be used to form the mixture density:

$$p_{w_T}(y_{T+1}|y_{1:T}) = \sum_{k=1}^{K} w_{T,k} p(y_{T+1}|y_{1:T}, M_k) \tag{1}$$

where the weight vector $w_T = (w_{T,1}, \dots, w_{T,K})'$ depends on data until time T and satisfies $\sum_{k=1}^{K} w_{T,k} = 1$ and $w_{T,1}, \dots, w_{T,K} \geq 0$. The naive equally-weighted model averaging (EWMA) method results when $w_{T,k} = 1/K$ for all k. By setting $w_{T,k} = p(M_k|y_{1:T})$ for all k, we obtain the formal Bayesian model averaging (BMA) method proposed by [7]. Assuming equal prior model probabilities, the kth posterior model probability (PMP) can be written as:

$$p(M_k|y_{1:T}) = \frac{p(y_{1:T}|M_k)}{\sum_{l=1}^{K} p(y_{1:T}|M_l)}.$$

In what follows, marginal likelihoods $p(y_{1:T}|M_k)$ are estimated by bridge sampling [8].

BMA presumes that the data generating process (DGP) belongs to $\mathcal{M}$. As this is questionable, we also consider a heuristic method that does not make this assumption; the optimal pooling (OP) method developed by [4, 5]. The OP weights are obtained by solving:

$$\max_{w_T} \sum_{t=t_0+1}^{T} \ln \left[\sum_{k=1}^{K} w_{T,k} p(y_t|y_{1:t-1}, M_k) \right]$$

$$\text{subject to} \sum_{k=1}^{K} w_{T,k} = 1 \text{ and } w_{T,1}, \dots, w_{T,K} \geq 0$$

where the objective function is the cumulative log score of (1) over $y_{t_0+1:T}$ given the training sample $y_{1:t_0}$.

3 Setting Up the Experiment

A forecasting competition is held between AR, LSTAR and RW models for a logistic transform of the monthly US unemployment rate and averages of these models. We assume a multivariate normal prior for autoregression coefficients in the AR and LSTAR models, an independent inverted gamma prior for σ^2 in each model and two independent normal priors for γ and c in the LSTAR model. We estimate the AR model with the Gibbs sampler, the LSTAR model with the Metropolis-within-Gibbs developed by [2]

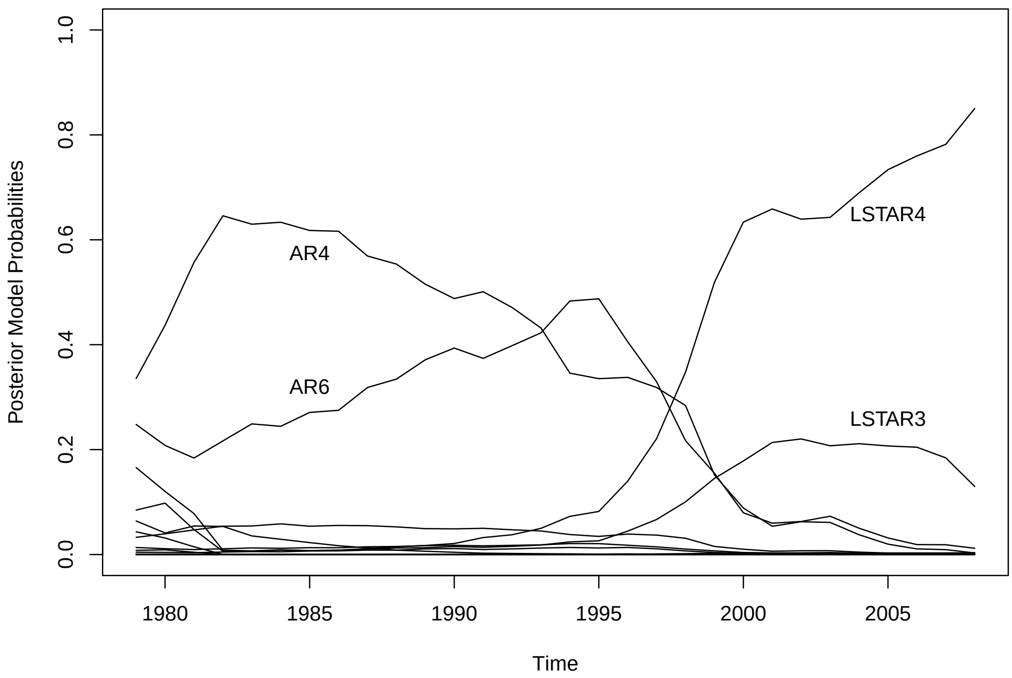

Figure 1: Evolution of PMPs over time for the AR(p) and LSTAR(p), $p = 1, \ldots, 8$. The PMPs are computed over expanding samples starting in 2:1949.

and use analytical results for the RW model.[1] Figure 1 enables us to determine the composition of the model averages; the four emerging models are retained for the BMA, OP and EWMA methods.[2] Finally, one-month ahead predictive densities are simulated from 1:1980 to 12:2009 using expanding estimation windows starting in 2:1949.

4 Results

Surprisingly, the real-time weights produced by BMA and OP exhibit a similar pattern: linearity is favored until roughly the middle of the forecasting period, while nonlinearity dominates afterward.

Predictive performance is evaluated with the Diebold-Mariano test [3], the efficiency test of West and McCracken [13], the log scoring rule and probability integral transforms (PITs). The statistical tests favor BMA and OP over the other models. Regarding the log scoring rule, BMA and OP perform well although they are outperformed by the LSTAR(4). Furthermore, it is difficult in all approaches to discriminate between BMA and OP. Finally, the PITs do not help for comparing predictive performance, but provide insights about misspecification issues.

[1] In the LSTAR, s_t is the lagged annual difference of the untransformed unemployment rate.

[2] As OP may sometimes attribute positive weights to inferior models, the RW model is also retained for this method.

References

[1] M. P. Clements and J. Smith. Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment. *Journal of Forecasting*, 19(4):255–276, 2000.

[2] P. J. Deschamps. Comparing Smooth Transition and Markov Switching Autoregressive Models of US Unemployment. *Journal of Applied Econometrics*, 23(4):435–462, 2008.

[3] F. X. Diebold and R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

[4] J. Geweke and G. Amisano. Optimal Prediction Pools. *Journal of Econometrics*, 164(1):130–141, 2011.

[5] J. Geweke and G. Amisano. Prediction with Misspecified Models. *American Economic Review*, 102(3):482–486, 2012.

[6] G. Koop and S. M. Potter. Dynamic Asymmetries in US Unemployment. *Journal of Business and Economic Statistics*, 17(3):298–312, 1999.

[7] E. E. Leamer. *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* Wiley, 1978.

[8] X.-L. Meng and W. H. Wong. Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration. *Statistica Sinica*, 6(4):831–860, 1996.

[9] A. L. Montgomery, V. Zarnowitz, R. S. Tsay, and G. C. Tiao. Forecasting the US Unemployment Rate. *Journal of the American Statistical Association*, 93(442):478–493, 1998.

[10] P. Rothman. Forecasting Asymmetric Unemployment Rates. *The Review of Economics and Statistics*, 80(1):164–168, 1998.

[11] T. Teräsvirta. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

[12] D. van Dijk, T. Teräsvirta, and P. H. Franses. Smooth Transition Autoregressive Models: A Survey of Recent Developments. *Econometric Reviews*, 21(1):1–47, 2002.

[13] K. D. West and M. W. McCracken. Regression-Based Tests of Predictive Ability. *International Economic Review*, 39(4):817–840, 1998.