Likelihood Tempering in Dynamic Model Averaging

Jan Reichl and Kamil Dedecius

Abstract We study the problem of online prediction with a set of candidate models using dynamic model averaging procedures. The standard assumptions of model averaging state that the set of admissible models contains the true one(s), and that these models are continuously updated by valid data. However, both these assumptions are often violated in practice. The models used for online tasks are often more or less misspecified and the data corrupted (which is, mathematically, a demonstration of the same problem). Both these factors negatively influence the Bayesian inference and the resulting predictions. In this paper, we propose to suppress these issues by extending the Bayesian update by a sort of likelihood tempering, moderating the impact of observed data to inference. The method is compared to the generic dynamic model averaging and to an alternative solution via sequential quasi-Bayesian mixture modeling.

Key words: Model averaging, model uncertainty, prediction, sequential estimation, tempered likelihood.

1 Introduction

In many real-world applications of the statistical control theory we are interested in online prediction of process outcomes, evaluated by an a priori specified process model. In practice, it is mostly assumed that the adopted process model is sufficiently close to the true observations-generating model. However, this (surprisingly still prevailing) assumption is often violated due to various reasons, e.g., different operational regimes, noise heteroskedasticity or even unknown but complicated distribution, imperfect physical characterization of the observed process etc. If there exists a class of potentially admissible candidate models, maybe with different input explanatory data and valid through different periods of time, the model switching and averaging procedures represent an appealing way around the model uncertainty problem [7, 16]. Model switching is mostly based on the assumption of mutually exclusive models. Their probabilities are sequentially assessed and the model with the highest probability is used for a corresponding period of time. Model averaging virtually removes this exclusivity via concurrent predictive-performance-based assessment of uncertainty about the candidate models.

Jan Reichl
Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic, e-mail: reichja3@fjfi.cvut.cz

Kamil Dedecius
Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic, e-mail: dedecius@utia.cas.cz

Although the model switching and averaging approaches admit that the adopted process models are not exact, they still ignore the possible misspecification issue in their inference. Indeed, even the best *available* models may be more or less misspecified, and provide only an approximation to the distribution of the observations. Although the consistency of the Bayesian inference guarantees the convergence of parameter estimates to the value minimizing the Kullback-Leibler divergence between the true distribution and its (imprecise) model [4], the results may be unacceptable for real applications. Among the typical demonstrations of such effects are negative estimates of strictly positive variables (fuel consumption, number of particles, object length etc.). Even worse, the standard Bayesian procedures are generally not robust to contamination by observations following other than the specified model, or to complete model misspecification [14].

This paper adheres to the dynamic model avaraging (DMA) realm, providing parallel online assessment of candidate models probabilities. If the modeling goal lies in evaluating the predicted values, the DMA-point prediction is represented by a convex combination of individual models' predictions, whose contributions are proportional to the candidate model probabilities. The DMA method was originally proposed for linear models by Raftery *et al.* [16] as an online extension of the static Bayesian model averaging [12, 17], and later formulated for sequential logistic regression by McCormick *et al.* Since then, DMA attained a significant focus, mostly in econometrics and finance, e.g. [2, 3, 6, 9, 10, 11] to name a few.

More specifically, we focus on the issue of assimilation of model-incompatible information (observations) into the prior distribution during the online estimation process, where the prior distributions involved in DMA are updated regardless of the candidate models probability. It leads to unreliable estimates and may degrade the prediction quality for a long time period. Although the invalid information may be gradually eliminated by means of forgetting procedures (see [5] for an extensive overview), a rational argument is to prevent it from entering the prior distribution. For this reason, we study the possibilities of online likelihood tempering during the Bayesian inference and prediction. Our method is inspired by the weighted likelihood [20], recently appearing also in the c-posterior approach [14] aiming at the same objective – the robustness to model misspecification. The approach sketched in this paper is the first step towards extending of the c-posteriors to online modeling. The determination of tempering factors is based on the actual predictive performance of the particular models. In particular, we study two possibilities: (i) a simplified approach based on the (dis)similarity of the most likely observation and the true one, and (ii) a model-oriented approach where the model weights and tempering factors are estimated by means of the quasi-Bayesian framework for online mixture inferrence [8, 18].

For completeness we remark that the likelihood tempering may be found in the Markov chain Monte Carlo as the MC^3 likelihood tempering method [1], however, the motivation is different.

The paper is organized as follows: Section 2 overviews the principles of the dynamic model averaging method. It also sheds some light on the studied problem of the connection between model uncertainty and observations assimilation. Section 3 is devoted to the proposed tempered sequential Bayesian update. Section 4 studies an alternative approach inspired by the quasi-Bayesian mixture modeling. The ongoing Section 5 illustrates the effect of tempering on a simulation example. Finally, Section 6 concludes the paper.

2 On-line Prediction with a Set of Admissible Models

In this section we describe the principles of the dynamic model averaging applied to a set of K admissible candidate models indexed by $k = 1, \ldots, K$. We consider discrete-time Bayesian modeling of a dynamic process with common observations y_t that are more or less determined by known possibly model-specific explanatory variables $x_{k,t}$, where $t = 0, 1, \ldots$ is the discrete time index. Assume, that some parametric models – probability density functions $p_k(y_t|x_{k,t}, \theta_k)$ parameterized by θ_k – are the admissible candidates for this purpose, and that (proper) prior distributions $\pi_k(\theta_k|x_{k,0:t-1}, y_{0:t-1})$ serve for their inference. The variables

$$x_{k,0:t-1} = \{x_{k,0}, \ldots, x_{k,t-1}\} \qquad \text{and} \qquad y_{0:t-1} = \{y_0, \ldots, y_{t-1}\}$$

express the statistical knowledge about θ_k up to time instant $t-1$, and $x_{k,0}$ and y_0 stand for pseudo-observations, expressing the initial prior knowledge before incorporation of the first observations. Let the modeling be performed under the uncertainty which model is true at the moment. The task is online prediction of y_t from $x_{k,t}$ and θ_k by means of model averaging.

2.1 Dynamic Model Averaging

As mentioned in the Introduction, the dynamic model averaging (DMA) methodology of Raftery *et al.* [16] extends Leamer's static Bayesian model averaging [12] to online problems, where the predictions are evaluated from sequentially acquired data.

The basic principle of DMA is that it assigns the candidate models $p_k(y_t|x_{k,t}, \theta_k)$ with probabilities (weights) $w_{k,t}$ taking values in the probabilistic $(K-1)$-simplex, that express the degree of evidence that these models are valid at the particular time instants. Recall, that the goal is the online prediction of the next observation y_t given known explanatory variables $x_{k,t}$. Each of the models provides its own point prediction, mostly represented by the expected value

$$\widehat{y}_{k,t} = \mathbb{E}\left[y_t|x_{k,0:t,y_{0:t-1}}\right] = \int y_t p_k(y_t|x_{0:t}, y_{0:t-1}) dy_t, \tag{1}$$

where

$$p_k(y_t|x_{0:t}, y_{0:t-1}) = \int p_k(y_t|x_t, \theta_k)\pi_k(\theta_k|x_{k,0:t-1}, y_{0:t-1}) d\theta_k \tag{2}$$

is the predictive distribution connected with the kth model $p_k(y_t|x_{k,t}, \theta)$. The integrations are over the spaces of θ_k and y_t, respectively. The DMA point prediction reflects the uncertainty about the particular models by taking their probabilities into account, that is, it averages over all the available predictions,

$$\widehat{y}_t = \sum_{k=1}^{K} w_{k,t-1}\widehat{y}_{k,t}. \tag{3}$$

After acquiring the observation y_t, the update of the probabilities $w_{k,t-1}$ reflects the predictive performance of the particular models,

$$w_{k,t} \propto w_{k,t-1} \cdot p_k(y_t|x_{k,0:t}, y_{0:t-1}). \tag{4}$$

In practice, the distribution of model weights may be considerably influenced by outdated information. This issue can be resolved by an artificial increase of the uncertainty about the weights, e.g., by exponential forgetting, flattening $w_{1,t-1}, \ldots, w_{K,t-1}$ by a factor $\alpha \in [0,1]$ as proposed by Peterka [15]. Furthermore, aberrant observations may get some $w_{k,t-1}$ too close to zero. This situation virtually eliminates the related models, as it is hard to recover from it. A workaround is to proceed with a stabilization additive constant during the weights update (4), e.g. $c = 10^{-3}/K$ proposed in [16]. The resulting equivalent of (4) then has the form

$$w_{k,t} \propto (w_{k,t-1}^{\alpha} + c) \cdot p_k(y_t|x_{k,0:t}, y_{0:t-1}). \tag{5}$$

The estimation of parameters $\theta_k, k = 1, \ldots, K$ is not influenced by DMA and has the form of the standard Bayes' theorem

$$\pi_k(\theta_k|x_{k,0:t}, y_{0:t}) \propto p_k(y_t|x_{k,t}, \theta_k)\pi_k(\theta_k|x_{k,0:t-1}, y_{0:t-1}). \tag{6}$$

We emphasize, that this is where the following question arises:

The dynamic model averaging (DMA) is used to assess the probability of several candidate models. Would it be possible and useful to take this probability into account in Equation (6)*?*

In other words, if one knows from $w_{k,t}$ that y_t are not well explained by $p_k(y_t|x_{k,t},\theta_k)$, why should one violate the estimates provided by $\pi_k(\theta_k|\cdot)$? For instance, if the reality switches between two models, their prior distributions are updated regardless of which of the models is currently valid.

Below, we propose to solve this issue by likelihood tempering.

3 Tempered Bayesian Update

Let us drop the model indices k in this section. The standard Bayesian consistency theory assumes that the there is a true observations-generating model $q(y_t|x_t)$ which is approximated by the statistician using a parametric model $p(y_t|x_t,\theta)$, ideally as close as possible. Under certain assumptions, the posterior estimates then converge to the value minimizing the Kullback-Leibler divergence of the two models,

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \mathbb{D}\left(q(y_t|x_t)\big|\big|p(y_t|x_t,\theta)\right),$$

where Θ is the parameter space, see, e.g., [4]. That is, the classical Bayesian way of thinking admits, that there is a possible disagreement between the true but unknown model and the employed (approximate) model, and relying on the consistency of the Bayesian posterior distribution of θ, it updates the prior distribution via the Bayes' theorem

$$\pi(\theta|x_{0:t},y_{0:t}) \propto p(y_t|x_t,\theta)\pi(\theta|x_{0:t-1},y_{0:t-1}). \tag{7}$$

However, from the Bayesian asymptotic theory it is well known that (7) requires certain assumptions to provide "reasonable" results. Naturally, if these assumptions are not satisfied, e.g., some observations are not explained by the model, the posterior estimates are inappropriately influenced (biased) and unusable for prediction. This effect is pronounced in DMA, where the standard Bayesian update is used, too.

We propose to solve this issue by a weighted variant of the Bayesian update, suppressing the influence of such observations similarly as in the Miller and Dunson's version for static estimation [14]. It consists of a step to increase the uncertainty about the model using a tempering (flattening) factor $\zeta_t \in [0,1]$,

$$\pi(\theta|x_{0:t},y_{0:t}) \propto \left[p(y_t|x_t,\theta)\right]^{\zeta_t}\pi(\theta|x_{0:t-1},y_{0:t-1}). \tag{8}$$

Miller and Dunson also propose a method for choosing a suitable value of ζ_t, however, it is not suitable for online cases.

A suboptimal solution of the first choice may be to base the factor on the predictive density and to compare the likelihood of the actually observed y_t with the expected $\widehat{y}_t$, i.e, the point estimate,

$$\zeta_t = \frac{p(y_t|x_{0:t},y_{0:t-1})}{p(\widehat{y}_t|x_{0:t},y_{0:t-1})}. \tag{9}$$

Naturally, it is possible to use the mode or other statistics in place of the mean value. Although this solution may lead to uncertainty underestimation, it could be easily counterbalanced by flattening of the posterior distribution, routinely used in engineering practice [5]. The predictive likelihood is analytically tractable in many practical cases, e.g. the linear regression models or Kalman filters, which allows for an easy computation of the factors in real time.

To summarize, the purpose of the proposed tempering update is to (i) penalize model misspecification, and (ii) to increase the robustness of estimation to contamination with other processes. The coefficient ζ_t can easily suppress the effect of unlikely observations (with respect to the predictive likelihood). If $\zeta_t \to 0$,

$$\pi(\theta|x_{0:t}, y_{0:t}) = \pi(\theta|x_{0:t-1}, y_{0:t-1}),$$

suppressing the influence of the extremely unlikely observation, while $\zeta_t \to 1$ recovers the standard Bayesian update. We suggest that this procedure should be used for estimation of parameters of models involved in DMA.

4 Mixture-Based Approach

Another possibility of model mixing[1] is to adopt the viewpoint that the class $\{p_k(y_t|x_{k,t}, \theta_k), k = 1, \ldots, K\}$ is a set of mixture components with weights $w_1, \ldots, w_K$, and to model these weights via the Dirichlet prior distribution whose hyperparameters $[\kappa_1, \ldots, \kappa_K]$ are updated in the sense of the quasi-Bayesian approach [18],

$$\pi(w|x_{1:K,0:t}, y_{0:t}) \propto \underbrace{\prod_{k=1}^{K} w_k^{\zeta_{k,t}}}_{\text{multinomial}} \underbrace{\prod_{k=1}^{K} w_k^{\kappa_{k,t-1}-1}}_{\text{Dirichlet prior}}, \tag{10}$$

where $\zeta_{k,t}$ is the estimate of the active component indicator,

$$\zeta_{k,t} \propto \frac{\kappa_{k,t-1}}{\sum_{l=1}^{K} \kappa_{l,t-1}} p_k(y_t|x_{k,0:t}, y_{0:t-1}).$$

Similarly to the Bayesian counterparts of the expectation-maximization algorithms, the quasi-Bayesian mixture estimation already assumes weighted Bayes' theorem for the estimation of component parameters,

$$\pi(\theta_k|x_{k,0:t}, y_{0:t}) \propto [p_k(y_t|x_{k,t}, \theta_k)]^{\zeta_{k,t}} \pi(\theta_k|x_{k,0:t-1}, y_{0:t-1}).$$

The analytical updates of Dirichlet prior hyperparameters (10) are given by

$$\kappa_{k,t} = \kappa_{k,t-1} + \zeta_{k,t}.$$

Finally, the prediction of the upcoming observation is similarly to DMA a convex combination of component predictions weighted by component weights,

$$\widehat{y}_t = \frac{1}{\kappa_{0,t}} \sum_{k=1}^{K} \kappa_{k,t} \widehat{y}_{k,t}, \qquad \text{where} \qquad \kappa_{0,t} = \sum_{k=1}^{K} \kappa_{k,t}. \tag{11}$$

The key differences between this approach and the weighted DMA proposed above are apparent from the equations. First and foremost, the component weights are modelled as static (or slowly varying if a forgetting procedure is used). In the authors' viewpoint, this is where the main drawback lies. In many practical situations, the models switch rather abruptly, which needs to be reflected by quick changes of their weights. However, the forgetting procedures can be effective only in the cases of slow variations [5], and under abrupt changes lead to significantly biased weights estimation. This will be demonstrated in the following section. For this reason, this section is rather conceptual and included for completeness as another model-based dynamic approach to model mixing.

[1] Proposed in personal communication by Dr. Kárný (Institute of Information Theory and Automation, Czech Academy of Sciences).

5 Simulation Results

In this section we present some simulation results comparing the studied strategies. The experiment is done on synthetic data generated by three models and contaminated with Laplacian noise. The performance is measured by the prediction mean squared error (MSE) and the mean absolute error (MAE).

The data set consists of 500 samples generated from the following normal models

$$
\begin{aligned}
y_t &= -0.2 - 0.15x_t^{(1)} + \varepsilon_t, && t = 150, \ldots, 300, \\
y_t &= 0.5 + 0.75x_t^{(2)} + \varepsilon_t, && t = 1, \ldots, 150 \text{ and } t = 300, \ldots, 400, \\
y_t &= 0.95x_t^{(3)} + \varepsilon_t, && t = 400, \ldots, 500,
\end{aligned}
$$

with regressors $x_t^{(i)} \sim \mathcal{N}(\mu_i, 1)$, where $\mu_1 = 0, \mu_2 = 2$ and $\mu_3 = 3$. The normal noise variable ε_t is drawn from $\mathcal{N}(0, 1.25^2)$.

Two different scenarios are studied:

1. *'True' model scenario* where the three models are used to generate data and no misspecification is present. Effectively, it is a true model switching scenario.
2. *Misspecification scenario* where the data is additionally contaminated by a heavy-tailed Laplacian noise, namely by 200 samples from zero-mean Laplace distribution with a scale 2. These samples are randomly added to the data.

The resulting data sets depicts Fig. 1, where the red crosses indicate the 'true' model scenario data, while the blue dots show Laplacian noise-contaminated data for the misspecified scenario.

The process y_t is modeled with three normal linear models

$$
y_t \big| x_t, \beta, \sigma^2 \sim \mathcal{N}\left(x^\intercal \beta, \sigma^2\right), \tag{12}
$$

where x_t represents 2-dimensional real regressors $x_t^{(1)}, \ldots, x_t^{(3)}$. Clearly, these models are not appropriate for the misspecification scenario with the Laplacian noise.

The prior placed on $\beta^{(i)}, \sigma^{2,(i)}$ is the normal-inverse gamma distribution in the compatible form conjugate to the exponential family form of the normal model (12) (more on this can be found, e.g., in [15]) with the same initialization for all considered models, namely with the sufficient statistic accumulating hyperparameter and the scalar degrees of freedom

$$
\xi_0^{(i)} = \mathrm{diag}(0.1, 0.01, 0.01) \quad \text{and} \quad \nu_0^{(i)} = 10,
$$

respectively. The posterior predictive distribution is in this setting the Student's distribution [15].

Three averaging strategies are studied:

1. Basic dynamic model averaging (DMA) of Raftery *et al.* [16],
2. Tempered dynamic model averagning (t-DMA) proposed in Section 3, and
3. Quasi-Bayesian approach (q-B) inspired by mixture modelling and discussed in Section 4.

The initial setting of all three strategies is identical. Namely, the initial model weights are uniform, the initial prior for weights in q-B is the flat Dirichlet distribution, symmetric and uniform over the related simplex.

The predictive performance is measured in terms of the mean squared error (MSE) and the mean absolute error (MAE). Consistently with the forecasting theory, the goal is to minimize these measures.

The results are summarized in Tab. 1 for the individual models without switching, and for all three studied averaging approaches. Both the misspecification scenario (MSE and MAE) and the 'true model' scenario (MSE true and MAE true) are shown. From the results one may conclude that the tempered DMA performs best (in terms of MSE and MAE). That is, the weighted Bayesian update (8) effectively suppresses the influence of data that are not explained by the model(s) in use. The classical DMA strategy performs a bit worse as expected. The quasi-Bayesian approach leads to a prediction quality inferior to both t-DMA and DMA. Anyway, it can provide results better than certain isolated models.

The time evolution of the models weights is depicted in Fig. 2 for both the original DMA and the proposed tempered version. The evolution of q-B is omitted for its inferior quality. Apparently, the initial learning period is very similar in both DMA and t-DMA, however, the model switching after $t = 150$ is much better reflected by t-DMA (the weight of the corresponding model is mostly close to 1). Another model switches are better detected by t-DMA too, although the weights are not so pronounced. One can conclude that t-DMA is significantly more sensitive than the basic DMA.[2]

To summarize, the experimental results show that the dynamic model averaging strategy performs well even in complicated conditions, where the noise properties are different from the assumed and the models differ from the true ones. The proposed tempering strategy leads to better results than the pure DMA ignoring the fact that the observations are not well described by the models and fully assimilating them into the prior distributions. Our experience confirms that if the true model is present, its weight is dominant.

Model/Strategy	MSE	MAE	MSE true	MAE true	$\widehat{\omega}_1$	$\widehat{\omega}_2$	$\widehat{\omega}_3$
p_1	7.59	1.92	5.13	1.75	-	-	-
p_2	5.44	1.47	2.61	1.22	-	-	-
p_3	5.22	1.37	2.51	1.13	-	-	-
DMA	4.86	1.29	2.14	1.08	0.21	0.41	0.36
t-DMA	4.47	1.14	1.77	0.91	0.32	0.38	0.30
q-B	6.11	1.35	3.21	1.16	0.17	0.37	0.44

Table 1 Prediction MSEs and MAEs and averaged models weights $\widehat{\omega}$ of all tested strategies (DMA, tempered DMA and the quasi-Bayesian approach) and single models without switching/averaging. For comparison, MSE true and MAE true denote error statistics computed with respect to the actual non-contaminated observation.

6 Conclusion

Applied statistical theory assumes that the models in use are relatively close to the true observations-generating models, and hence that the results (predictions, estimates) are close to the true values. However, the real-world phenomena are often rather roughly approximated by models. The standard dynamic model averaging provides a way around this issue by means of a concurrent assessment of models uncertainty, and by averaging over the results taking this uncertainty into account. However, it still neglects it at the level of the Bayesian update. In this paper, we focus specifically on this issue and propose to use a weighted – tempered – version of the Bayes' theorem, suppressing the impact of unlikely observations to the inference. The simulation example demonstrates that it provides an improvement of estimation quality.

Acknowledgement

The project is supported by the Czech Science Foundation, project number 14–06678P. The authors thank to Dr. M. Kárný and Dr. L. Jirsa for the discussions about the model-based approach.

[2] A thorough sensitivity analysis is postponed to further research.

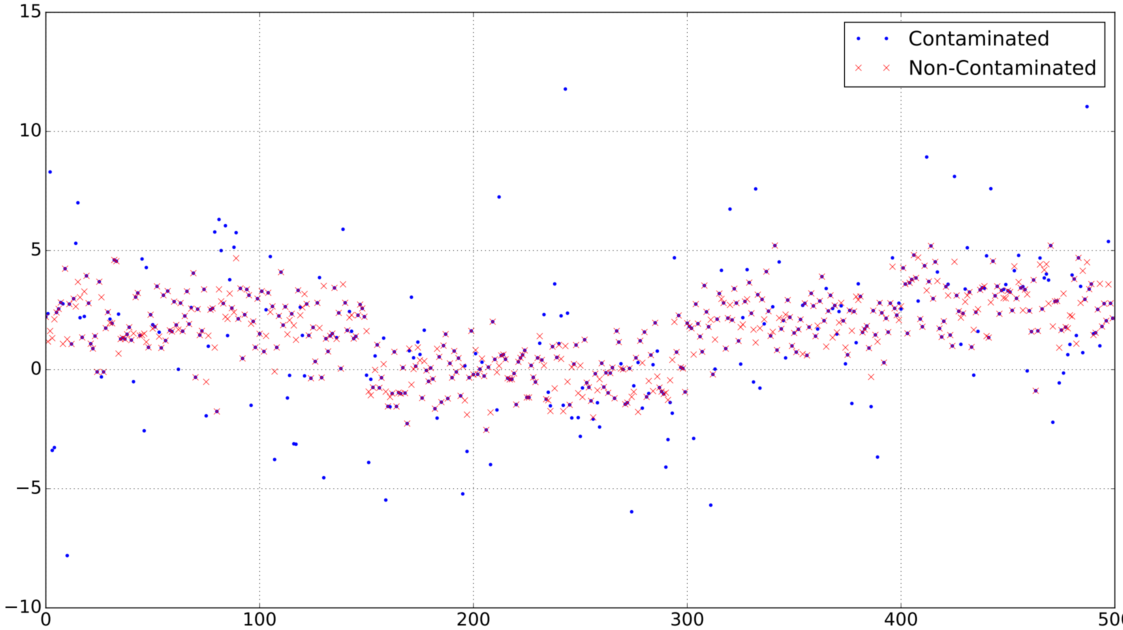

Fig. 1 Comparison of contaminated and non-contaminated data set.

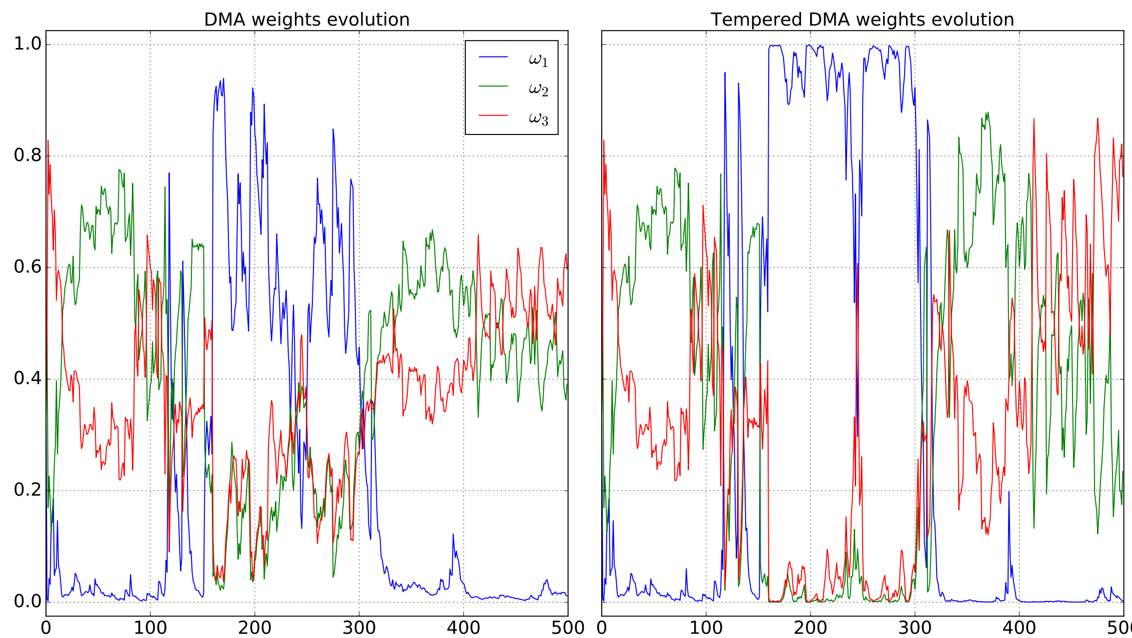

Fig. 2 Time evolution of model weights for the DMA and tempered-DMA. The correspondence between models and line types is the same in both plots.

References

1. Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F.: Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics **20**(3), 407–415 (2004)
2. Aye G., Gupta, R., Hammoudeh, S., Kim W.J.: Forecasting the price of gold using dynamic model averaging. International Review of Financial Analysis, **41**, 257–266 (2015)
3. Bork, L., Møller, S.V.: Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection, International Journal of Forecasting, **31**(1), 63–78 (2015)
4. Bunke, O., Milhaud, X.: Asymptotic behavior of Bayes estimates under possibly incorrect models. The Annals of Statistics **26**(2), 617–644 (1998).
5. Dedecius, K., Nagy, I., Kárný, M.: Parameter tracking with partial forgetting method. International Journal of Adaptive Control and Signal Processing **26**(1), 1–12 (2012)
6. Del Negro, M., Hasegawa, R.B., Schorfheide, F.: Dynamic prediction pools: An investigation of financial frictions and forecasting performance. Journal of Econometrics, **192**(2), 391–405 (2016)
7. Frühwirth-Schnatter, S.: Finite Mixture and Markov Switching Models. Springer (2006)
8. Kárný, M., Kadlec, J., Sutanto, E.L.: Quasi-Bayes estimation applied to normal mixture. In: Proceedings of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, pp. 77–82. Praha (1998)
9. Koop, G., Korobilis, D.: Forecasting inflation using dynamic model averaging. International Economic Review, **53**(3), 867–886 (2012).
10. Koop, G., Korobilis, D.: A new index of financial conditions. European Economic Review, **71**, 101–116 (2014)
11. Koop, G., Tole, L.: Forecasting the European carbon market. Journal of the Royal Statistical Society: Series A (Statistics in Society), **167**(3), pp. 723–741 (2013).Journal of the Royal Statistical Society: Series A (Statistics in Society)
12. Leamer, E.E.: Specification searches: Ad hoc inference with nonexperimental data. Wiley, New York (1978)
13. McCormick, T., Raftery, A., Madigan, D., Burd, R.: Dynamic logistic regression and dynamic model averaging for binary classification. Biometrics **68**(1), 23–30 (2012).
14. Miller, J.W., Dunson, D.B.: Robust Bayesian inference via coarsening pp. 1–31 (2015). URL http://arxiv.org/abs/1506.06101
15. Peterka, V.: Bayesian approach to system identification. In: P. Eykhoff (ed.) Trends and Progress in System Identification, pp. 239–304. Pergamon Press, Oxford, U.K. (1981)
16. Raftery, A.E., Kárný, M., Ettler, P.: Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. Technometrics **52**(1), 52–66 (2010)
17. Raftery, A.E., Madigan, D., Hoeting, J.A.: Bayesian model averaging for linear regression models. JASA, **92**(437), 179–191 (1997)
18. Titterington, D.M., Smith, A.F.M., Makov, U.E.: Statistical analysis of finite mixture distributions. John Wiley (1985)
19. Raiffa, H., Schlaifer, R.: Applied Statistical Decision Theory (Harvard Business School Publications). Harvard University Press (1961).
20. Wang, X.: Approximating Bayesian inference by weighted likelihood. Cannadian Journal of Statistics, **34**(2), 279–298 (2006)