Adaptive learning from model space*

Jan Prüser[a,b]

[a]University of Duisburg-Essen† [b]Ruhr Graduate School in Economics‡

June 18, 2018

Dynamic model averaging (DMA) is used extensively for the purpose of economic forecasting. This study extends the framework of DMA by introducing adaptive learning from model space. In the conventional DMA framework all models are estimated independently and hence the information of the other models is left unexploited. In order to exploit the information in the estimation of the individual time-varying parameter models, this paper proposes to not only average over the forecasts but, in addition, to also dynamically average over the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior, which is then used in the upcoming period as the prior for each of the individual models. The relevance of this extension is illustrated in three empirical examples involving forecasting US inflation, US consumption expenditures and forecasting of five major US exchange rate returns. In all applications adaptive learning from model space delivers improvements in out-of-sample forecasting performance.

Keywords: Forecasting, Fat Data, Variable Selection, Model Change
JEL classification: C11, C32, C53, E37

*I thank Christoph Hanck and Alexander Schlösser for valuable comments.

†Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany.

‡RWI - Leibniz Institute for Economic Research, Hohenzollernstrasse 1-3, D-45128 Essen, Germany, e-mail address: `jan.prueser@rgs-econ.de`.

1. Introduction

Forecasting in economics is challenging for three major reasons. First, the existence of many potential predictors can result in a huge number of potential models. While regressions with many predictors may overfit, small models may miss important predictors. This leads to the need for model selection strategies. Second, a useful forecasting model may change over time. For instance, some variables may predict well in recessions while others may predict well in expansions or the set of relevant predictors may change between certain events such as the Great Moderation. This further complicates the statistical problem as a researcher needs to select one model in each period. Third, in case of parameter change the marginal effect of predictors may change over time. However, modeling such change will increase the risk of overfitting the data, resulting in poor out-of-sample predictions. Recently, a growing literature addresses these points by using dynamic model averaging (DMA), proposed by Raftery et al. (2010). Koop and Korobilis (2012) introduce DMA to the economic literature by forecasting inflation. They find a favorable forecasting performance of DMA over simple benchmark regressions and more sophisticated approaches. Studies that use DMA to forecast a variety of different economic time series include: Buncic and Moretto (2015), Drachal (2016) and Naser (2016) forecasting commodities, Bruyn et al. (2015), Beckmann and Schüssler (2016) and Byrne et al. (2018) forecasting exchange rates, Liu et al. (2015) forecasting stock returns, Gupta et al. (2014) forecasting foreign exchange reserves, Bork and Moller (2015), Risse and Kern (2016) and Wei and Cao (2017) forecasting house price growth, Aye et al. (2015) and Baur et al. (2016) forecasting gold prices, Koop and Korobilis (2011) and Filippo (2015) forecasting inflation and Wang et al. (2016) and Liu et al. (2017) forecasting realized volatility.

While conventional DMA is well established in the economic literature, the aim of this paper is to extend this framework by introducing adaptive learning from model space (ALM). The conventional DMA approach consists of independently estimating K different time-varying parameters (TVP) models. In order to combine the different models, their individual forecasts are weighted by time-varying inclusion probabilities. The time-varying inclusion probabilities depend on the most recent forecasting performance of each model and allow that the weight placed on each model may change over time. However, as each model is estimated independently and the information provided by the other models and the information provided from the time-varying inclusion probabilities is left unexploited in the process. In order to exploit the information in the estimation of the individual TVP models, this paper proposes to not only average over the forecasts but, in addition, to also dynamically average over the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior in each period. By doing so, the information of all model posteriors is summarized in one single posterior, which is then used in the following period as the prior for each of the individual models. This is attractive because it is often argued that pooling information is optimal relative to pooling forecasts, as the latter introduces an efficiency loss, see Timmermann (2006). For instance it may be the case that at the beginning of the sample most weight is placed on parsimonious models while later in the sample more weight is place on models with a larger set of predictors. However, these models cannot benefit from this information when estimated independently. In contrast, when averaging over the time-varying parameters some of the variables in the larger models may be shrunk to zero at the beginning of the sample by exploiting the information that they were not relevant at this time.

The relevance of this extension is illustrated in three empirical applications. In the first application, both conventional DMA and ALM are used to forecast US inflation one quarter and one year ahead. Under different settings, ALM compares favorably to conventional DMA. The second application considers forecasting nominal and real US consumption expenditures one quarter and one year ahead. For nominal and real consumption expenditures ALM outperforms conventional DMA. Finally, the third application forecasts five major US end-of-month (log) exchange rate returns one month and one year ahead. It turns out that ALM delivers more precise forecasts than the conventional DMA for all five countries. The finding that ALM yields improvements in out-of-sample forecasting holds in particular for the long horizon in all three applications.

The remainder of this paper is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 presents the empirical findings and the last section concludes.

2. Econometric Framework

2.1. Baseline Dynamic Model Averaging

Consider a set of time-varying-parameters (TVP) models M_k, $k = 1, \ldots, K$, which can be written as

$$y_t = \boldsymbol{z}_t^{(k)} \boldsymbol{\theta}_t^{(k)} + \epsilon_t^{(k)}, \tag{1}$$

$$\boldsymbol{\theta}_t^{(k)} = \boldsymbol{\theta}_{t-1}^{(k)} + \boldsymbol{\eta}_t^{(k)}, \tag{2}$$

where $\epsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\boldsymbol{\eta}_t^{(k)} \sim N(\boldsymbol{0}, \boldsymbol{Q}_t^{(k)})$. The predictor vector $\boldsymbol{z}_t^{(k)}$ for each model can be of different dimension and the predictor set of the different models net not to overlap. Let $L_t = k$ if the process is modelled by model M_k at time t. Conditioning on $L_t = k$, the state vector $\boldsymbol{\theta}_t^{(k)}$ of each model can be estimated independently using the Kalman filter. Assuming that $\boldsymbol{\theta}_{t-1}^{(k)}|L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)}, \boldsymbol{\Sigma}_{t-1|t-1}^{(k)})$, Kalman filtering proceeds using

$$\boldsymbol{\theta}_t^{(k)}|L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}, \boldsymbol{\Sigma}_{t|t-1}^{(k)}), \tag{3}$$

where

$$\hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} = \hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)} \tag{4}$$

and

$$\boldsymbol{\Sigma}_{t|t-1}^{(k)} = \boldsymbol{\Sigma}_{t-1|t-1}^{(k)} + \boldsymbol{Q}_t^{(k)}. \tag{5}$$

Followed by the updating equations to complete the estimation

$$\boldsymbol{\theta}_t^{(k)}|L_t = k, \boldsymbol{y}^t \sim N(\hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)}), \tag{6}$$

where

$$\hat{\boldsymbol{\theta}}_{t|t}^{(k)} = \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} + \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)} (H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)})^{-1} (y_t - \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}) \tag{7}$$

and

$$\boldsymbol{\Sigma}_{t|t}^{(k)} = \boldsymbol{\Sigma}_{t|t-1}^{(k)} - \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)} (H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t^{\prime(k)})^{-1} \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)}. \tag{8}$$

In order to run the Kalman filter, one needs to know the variance $H_t^{(k)}$ of the observation equation and the covariance matrix $\boldsymbol{Q}_t^{(k)}$ of the transition equation. Estimating or simulating $H_t^{(k)}$ and $\boldsymbol{Q}_t^{(k)}$ running MCMC methods for each model would be computationally demanding. Therefore, it is convenient to use approximations that allow to estimate each model with only one iteration of the Kalman filter. The covariance matrix $\boldsymbol{Q}_t^{(k)}$ appears in the Kalman Filter only in equation (5). Following Raftery et al. (2010), equation (5) is replaced by

$$\boldsymbol{\Sigma}_{t|t-1}^{(k)} = \frac{1}{\lambda} \boldsymbol{\Sigma}_{t-1|t-1}^{(k)} \tag{9}$$

or, equivalently, $\boldsymbol{Q}_t^{(k)} = (1 - \lambda^{-1}) \boldsymbol{\Sigma}_{t-1|t-1}^{(k)}$, where λ is called the forgetting factor with $0 < \lambda \leq 1$. Forgetting factor approaches have a long tradition in the state space literature and a detailed motivation is given , e.g., by Fagin (1964) and Jazwinsky (1970). The forgetting factor implies that observations which are lagged by i periods receive the weight λ^i. The idea is similar to applying a rolling window regression with a window size of $\frac{1}{1-\lambda}$. Typically the value for λ is set close to one, in order to favor a gradual evolution of coefficients. Raftery et al. (2010) set $\lambda = 0.99$. For quarterly data, this means that observations from five years ago receive around 80% as much weight as observation of the last period. Setting $\lambda = 0.95$ implies that observations five years ago receive only 35% as much weight as the observation of the last period and would allow for higher degrees of parameter change. This suggests that the range of plausible values should be close to one. In section 2.3 a way to estimate λ over a small grid of values is discussed. With this simplification there is no need to estimate $\boldsymbol{Q}_t^{(k)}$ anymore. To estimate each model with only one Kalman filter iteration requires a method for estimating $H_t^{(k)}$. Following Koop and Korobilis (2012), $H_t^{(k)}$ is estimated by using an Exponentially Weighted Moving Average (EWMA)

$$H_t^{(k)} = \kappa H_{t-1}^{(k)} + (1-\kappa)(y_t - \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)})^2, \tag{10}$$

with $0 < \kappa \leq 1$. This estimator is a weighted average of $H_{t-1}^{(k)}$ and the squared residuals at time t, with κ a decay factor, similar to the forgetting factor λ, with effective window size $\frac{\kappa}{2} - 1$. Therefore, the value for κ is also typically set close to one. RiskMetrics (1996) set $\kappa = 0.97$ for monthly data and Koop and Korobilis (2012) set $\kappa = 0.98$ for quarterly data. See RiskMetrics (1996) for general properties of the EWMA estimator. After having replaced $\boldsymbol{Q}_t^{(k)}$ and $H_t^{(k)}$ with equations (9) and (10), all results are available in closed form and only one iteration of the Kalman filter is required for the estimation of each model. As a next step, a way to combine the models is needed. Raftery et al. (2010) propose to calculate time-varying model probabilities by using the following model prediction equation with forgetting factor α:

$$\pi_{k|t-1} = \frac{\pi_{k|t-1}^{\alpha}}{\sum_{l=1}^{K} \pi_{l|t-1}^{\alpha}} \tag{11}$$

and a model updating equation

$$\pi_{k|t} = \frac{\pi_{k|t-1} p_k(y_t|\boldsymbol{y}^{t-1})}{\sum_{l=1}^{K} \pi_{l|t-1} p_l(y_t|\boldsymbol{y}^{t-1})}, \tag{12}$$

where p_k denotes the predictive likelihood of model k. The predictive likelihood is a measure of forecasting performance and is defined as the predictive density evaluated at the actual outcome y_t. The predictive density is given by

$$y_t|L_{t-1} = k, \boldsymbol{y}^{t-1} \sim N(\boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)}, H_t^{(k)} + \boldsymbol{z}_t^{(k)} \boldsymbol{\Sigma}_{t|t-1}^{(k)} \boldsymbol{z}_t'^{(k)}). \tag{13}$$

In order to understand the role of the forgetting factor α, write$\pi_{t|t-1,k}$ as

$$\pi_{k|t-1} \propto \prod_{i=1}^{t-1} [p_k(y_{t-i}|y^{t-i-1})]^{\alpha^i}. \tag{14}$$

Thus, the model probabilities change over time according to the forecasting performance (measured by the predictive likelihood) in the recent past of each model. The forgetting factor α discounts the past forecasting performance in the same fashion as the forgetting factor λ and therefore controls the frequency of model change. As a special case, $\alpha = 1$ corresponds to conventional model averaging using the marginal likelihood. Hence, similar considerations as for λ apply and suggest to set α close to one. Section 2.3 discusses a way to estimate α. Now, given the model probabilities we can forecast using dynamic model averaging (DMA) via

$$\hat{y}_t^{DMA} = \sum_{l=1}^{K} \pi_{k|t-1} \boldsymbol{z}_t^{(k)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k)} \tag{15}$$

or using dynamic models selection (DMS) as

$$\hat{y}_t^{DMS} = \boldsymbol{z}_t^{(k^*)} \hat{\boldsymbol{\theta}}_{t|t-1}^{(k^*)}, \tag{16}$$

where k^* refers to the model with the maximum model probability at time $t-1$. Thus, each model is estimated independently using the Kalman filter and then either the forecast of each individual model is weighted by its probability at period $t-1$ or the forecast of one model is selected with the highest probability at time $t-1$.

2.2. Adaptive learning from model space

The aim of this paper is to go one step further by not only performing DMA over the individual forecast of each model but also doing DMA over the time-varying model parameters $\boldsymbol{\theta}_t^{(k)}$. Using conventional DMA, each model is estimated independently and cannot exploit the information in the other models. By averaging over $\boldsymbol{\theta}_t^{(k)}$ each model is not estimated independently anymore and uses the information of all other $K-1$ models. This is attractive because it may be the case that at the beginning of the sample most weight is placed on parsimonious models while later in the sample more weight is placed on models with a larger set of predictors. However, these models cannot benefit from this information when estimated independently. In contrast, when averaging over $\boldsymbol{\theta}_t^{(k)}$, some of the variables in the larger models may be shrunk to zero at the beginning of the sample by exploiting the information that they were not relevant at this time.

In order to average over $\boldsymbol{\theta}_t^{(k)}$ at each period t, a single Gaussian $q(\boldsymbol{\theta}_t) = N(\boldsymbol{\theta}_t | \overline{\boldsymbol{\theta}}_{t|t}, \overline{\boldsymbol{\Sigma}}_{t|t})$ is used to approximate a mixture of Gaussians $p(\boldsymbol{\theta}_t) = \sum_{k=1}^{K} \pi_{k|t} N(\boldsymbol{\theta}_t | \hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)})$.[1] The Kullback-Leibler divergence (KL) is a measure of the dissimilarity between two distribu-

[1]Note that the vector $\boldsymbol{z}_t^{(k)}$ may include a different set of variables for each model k. Hence, $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ may also correspond to a different set of variables. In order to account for this, zeros can be placed in the corresponding elements of $\boldsymbol{z}_t^{(k)}$ in case certain variables are not included.

tions. Therefore, the two moments $\overline{\boldsymbol{\theta}}_{t|t}$ and $\overline{\boldsymbol{\Sigma}}_{t|t}$ can be determined by minimizing the Kullback-Leibler divergence between $q(\boldsymbol{\theta}_t)$ and $p(\boldsymbol{\theta}_t)$ with respect to $\overline{\boldsymbol{\theta}}_{t|t}$ and $\overline{\boldsymbol{\Sigma}}_{t|t}$. The minimization problem is given by $q = \arg\min_q \mathrm{KL}(q||p)$ and the solution to this problem is given by

$$\overline{\boldsymbol{\theta}}_{t|t} = \sum_{k=1}^{K} \pi_{k|t} \hat{\boldsymbol{\theta}}_{t|t}^{(k)}, \tag{17}$$

$$\overline{\boldsymbol{\Sigma}}_{t|t} = \sum_{k=1}^{K} \pi_{k|t} (\boldsymbol{\Sigma}_{t|t}^{(k)} + (\hat{\boldsymbol{\theta}}_{t|t}^{(k)} - \overline{\boldsymbol{\theta}}_{t|t})(\hat{\boldsymbol{\theta}}_{t|t}^{(k)} - \overline{\boldsymbol{\theta}}_{t|t})'). \tag{18}$$

In the graphical model literature, this is called weak marginalization, as it preserves the first two moments, see Lauritzen (1992). The center of the distribution $q(\boldsymbol{\theta}_t)$ is just the weighted average of the mean $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ from all individual models. Thus, models with a higher posterior probability receive more weight. The total variance $\boldsymbol{\Sigma}_{t|t}^{(k)}$ arises from two sources of variability. The first source is the weighted average of the covariance-matrix $\boldsymbol{\Sigma}_{t|t}^{(k)}$ of the individual models and reflects the uncertainty about $\boldsymbol{\theta}_t$ which comes from the estimation of the individual models. And the second source is the weighted average of the squared difference between $\hat{\boldsymbol{\theta}}_{t|t}^{(k)}$ and $\overline{\boldsymbol{\theta}}_{t|t}$ and reflects the uncertainty through the heterogeneity between the different models. In order to complete estimation, the posterior $q(\boldsymbol{\theta}_t)$ serves as a prior for each model in the upcoming period and $\hat{\boldsymbol{\theta}}_{t-1|t-1}^{(k)}$ and therefore $\boldsymbol{\Sigma}_{t-1|t-1}^{(k)}$ in equation (4) and (5) are replaced by $\overline{\boldsymbol{\theta}}_{t-1|t-1}$ and $\overline{\boldsymbol{\Sigma}}_{t-1|t-1}$. Thus, the time-varying parameter vector $\boldsymbol{\theta}_t$ is estimated by exploiting the information of all K models and its elements are shrunken towards the parameters of models that receive a higher weight. For example, they can be shrunk towards zeros, if models that include the corresponding variable receive only little weight or they may be shrunken towards the value they have in models which receive a high weight.

2.3. Estimation of hyperparamter

In order to estimate a model, one has to determine the values of the hyperparameters λ, κ and α. Previous consideration suggests to set them close to one. Furthermore, the two hyperparamters λ and κ are estimated over a grid of values by treating different values as different models, i.e. by setting $\lambda = \lambda^{(k)}$ and $\kappa = \kappa^{(k)}$. Thus, if models discounting past data more strongly yield a better forecasting performance (measured by the predictive likelihood) in the recent past (which is controlled by α) a higher weight is placed on them. However, it is not possible to estimate the forgetting factor α in this fashion. Fortunately, Beckmann and Schüssler (2016) provide a way to integrate (sum) over a grid of values $\alpha_v \in (\alpha_1, \alpha_2, \ldots, \alpha_a)$ by replacing equation (11) and (12) with

$$\pi_{k|t-1} = \sum_{v=1}^{a} \pi_{k|t-1,\alpha_v} p(\alpha_v|I_{t-1}), \tag{19}$$

where $\pi_{k|t-1,\alpha_v} = \frac{\pi_{k|t}^{\alpha_v}}{\sum_{l=1}^{K} \pi_{l|t}^{\alpha_v}}$ and

$$\pi_{k|t} = \sum_{v=1}^{a} \frac{p_k(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_v}}{\sum_{l=1}^{K} p_l(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_v}} p(\alpha_v|I_t). \tag{20}$$

The posterior at time t of a particular grid point of the forgetting factor α is given by

$$p(\alpha_z|I_t) = \frac{\sum_{k=1}^{K} p_k(y_t|\boldsymbol{y}^{t-1})\pi_{k|t-1,\alpha_z} p(\alpha_z|I_{t-1})}{\sum_{v=1}^{a}\sum_{l=1}^{K} p_l(y_t|\boldsymbol{y}^{t-1})\pi_{l|t-1,\alpha_v} p(\alpha_v|I_{t-1})}. \tag{21}$$

3. Empirical Applications

3.1. Forecasting Inflation

This section considers one quarter and one year ahead forecasts for core inflation as measured by the Personal Consumption Expenditure (PCE) deflator. A standard set of variables is considered as potential predictors, reflecting the major theoretical explanations of inflation as well as variables which have been found to be useful in forecasting inflation in other studies. Potential predictors are the percentage change in the Dow Jones Industrial Average, the percentage change in employment, the log of housing starts, University of Michigan survey of inflation expectations, the percentage change in the money supply (M1), the percentage change of Spot Crude Oil Price (WTI), the change in the Institute of Supply Management index (Manufacturing), the percentage change in real personal consumption expenditures, the percentage change in real GDP, the percentage change in real Gross Private Domestic Investment (Residential), the spread between the ten year and three month Treasury bill, the three month Treasury bill and the unemployment rate. In addition, following Koop and Korobilis (2012), each model contains one intercept and two lags of inflation. All variables used are in quarterly frequency, seasonally adjusted and are obtained from the FRED database of the Federal Reserve Bank of St. Louis. The data are observed for the period 1978Q2 to 2016Q3 and the period from 1992Q1 to 2016Q3 is used to evaluate the out-of-sample forecast performance.

Based on this set of predictors, the performance of conventional DMA and ALM is compared. Furthermore, the performance for different choices of the forgetting factors is investigated. Koop and Korobilis (2012) set $\kappa = 0.98$ and focus on a modest range for the forgetting factors, i.e. $\alpha, \lambda \in (0.95, 0, 99)$, which they find to deliver a favorable forecasting performance over simple benchmark regressions and more sophisticated approaches. Thus, this set of values is used to forecast inflation. In order to not just rely on one value for the whole period, this section also considers estimating $\alpha \in (0.95, 0.99)$ and $\lambda \in (0.95, 0.99)$ dynamically. This allows the framework to switch between a gradual change in both coefficients and models ($\alpha = \lambda = 0.99$), a more rapid change in both coefficients and models ($\alpha = \lambda = 0.95$) or a mix of the two over time.

Table A.1 contains the results for the one quarter and one year ahead forecasting performance in terms of the root mean squared forecast error (RMSFE) and in terms of the mean absolute forecast error (MAFE). In addition, it shows the results of the Diebold-Marion test (DM-test) proposed by Diebold and Marion (1995) in order to investigate whether the forecasting errors of the conventional DMA approach differ statistically significant from those obtained from the ALM approach. ALM and conventional DMA deliver similar forecasting errors in all settings for one quarter ahead inflation. But for one year ahead ALM delivers statistical significantly smaller forecasting errors. Furthermore, the results show that allowing for a more rapid change between models and in coefficients yields better predictions. However, estimating the forgetting factors seems to be a useful strategy in order to avoid poor forecasts due to a poor selection of values for the forgetting factors. Figure B.1 compares the cumulative sum of the absolute forecast error over time of ALM and conventional DMA and DMS. This allows to assess the forecasting performance over time. The absolute forecast errors grow roughly linearly over time for both horizons and all three approaches. Only after the financial crisis a big jump can be observed for all setups. Hence, one can conclude that the forecast errors are fairly stable over time. Furthermore, it can be seen that the one year cumulative sum of absolute forecast errors after the financial crisis is lower for ALM compared to the conventional approach.

3.2. Forecasting Consumption Expenditures

This section compares the forecasting performance of ALM and conventional DMA for nominal and real consumption expenditures. As potential predictors, the percentage change in the Dow Jones Industrial Average, the percentage change in employment, the log of housing starts, the percentage change in the money supply (M1), University of Michigan survey of inflation expectations, University of Michigan survey of consumer sentiment, the spread between the ten year and three month Treasury bill, the three month Treasury bill, the unemployment rate, the inflation rate and (real) disposable income are used. In addition, each model contains one intercept and two lags of consumption expenditures and the same grid of forgetting factors is used for the estimation as in the inflation application. All variables used are quarterly, seasonally adjusted and are obtained from the FRED database of the Federal Reserve Bank of St. Louis. The data are observed for the period 1978Q2 to 2016Q3 and the period from 1992Q1 to 2016Q3 is used to evaluate the out-of-sample forecast performance.

The results turn out to be similar to the ones obtained for inflation. Table A.2 shows that ALM and conventional DMA deliver similar forecasting errors for one quarter ahead consumption expenditures and for one year ahead, ALM delivers smaller forecasting errors. Moreover the DM test reveals that the difference between the squared forecast errors is statistically significant for one year ahead predictions. Figure B.2 shows a similar pattern for the cumulative sum of the absolute forecast errors for the consumption expenditures compared to these obtained for inflation. Again, the absolute forecast errors grow roughly linearly over time for both horizons and all three approaches with the exception of the period after the financial crisis. And again, while the cumulative sum of the absolute forecast errors is similar for the short horizon, for the long horizon the cumulative sum of the absolute forecast errors obtained from the ALM approach are below the ones obtained from conventional DMA for the entire period.

3.3. Forecasting Exchange Rates

This section considers one month and one year ahead forecasting of five major US end-of-month (log) exchange rate returns. The five different countries are Canada, Denmark, the United Kingdom, Japan and Sweden. The monthly data range from 1975M2 to 2017M4 and the forecasting results are obtained after a trainingperiod of 50 months. Four regressors, based on economic theory, in addition to an intercept are considered as potential predictors. The first regressor is based on the uncovered interest parity condition (UIP) and is defined as

$$UIP_t = i_t - i_t^*, \tag{22}$$

where i_t denotes the nominal interest rate and i_t^* is the foreign nominal interest rate (measured by the money market rate and obtained from the IFS database). The second regressor is based on the deviation from the purchasing power parity condition (PPP) and is defined as

$$PPP_t = p_t - p_t^* - s_t, \tag{23}$$

where p_t denotes the log of the domestic price level (measured as the consumer price index and obtained from the FRED database), p_t^* the log of foreign price level and s_t denotes the log of the nominal exchange rate (measured as end-of-period exchange rates and obtained from the FRED database). The third regressor is based on the asymmetric Taylor (1993) rule (ART) and is defined as

$$ATR_t = 1.5(\pi_t - \pi_t^*) + 0.1(g_t - g_t^*) + 0.1(s_t + p_t^* - p_t), \tag{24}$$

where π_t is the domestic inflation rate, π_t^* is the foreign inflation rate, g_t is the domestic output gap and g_t^* is the foreign output gap. The output gap is measured as the deviation of real output (measured by the industrial production and obtained from the FRED database) from an estimate of potential output calculated using the Hodrick and Prescott (1996) filter. The parameter values (1.5, 0.1, 0.1) are a standard choice in the literature, due to Molodtsova and Papell (2009). The last regressor is based on the deviation form the monetary fundamentals (MF) and is defined as

$$MF_t = (m_t - m_t^*) - (prod_t - prod_t^*) - s_t, \tag{25}$$

where m_t is the log of the domestic money supply (measured as M1 if available and otherwise as M3 and obtained from the OECD database), m_t^* is the log of the foreign money supply and $prod_t^{(*)}$ is the log of the domestic (foreign) industrial production.

For the estimation of the two forgetting factors α and λ a wider range as before is considered, i.e. $\alpha, \lambda \in (0.80, 0.90, 0.95, 0.99, 1)$, which in one extreme nests the special case of constant parameter models ($\lambda = 1$) and no model change ($\alpha = 1$) and in the other extreme allows for a very rapid change in both coefficients and models ($\alpha = \lambda = 0.80$). The wider grid allows the predictive information in macroeconomic fundamentals for exchange rate returns to change fast over time, as suggested by Beckmann and Schüssler (2016). For the decay factor κ a tight grid around the value 0.96 (which is recommended by RiskMetrics (1996) for monthly data), i.e. $\kappa \in (0.95, 0.96, 0.97, 0.98, 0.99)$, is considered.

Table A.3 displays the results for the one month and one year ahead forecasting performance. The results show a clear pattern. It turns out that ALM delivers a better forecasting performance for all five countries than conventional DMA and the difference in forecasting performance is statistically significant for one year ahead predictions in all cases and in most cases for one month ahead predictions. Figure B.3 compares the cumulative sum of the absolute forecast error over time of ALM and conventional DMA and DMS. The cumulative sum of absolute forecast errors for one month grows roughly linear over time. In contrast, the cumulative sum of the absolute forecast error of one year exhibits some jumps over time. This shows that the absolute forecast errors of one year ahead are less stable than for one month ahead. Moreover, while for the short horizon the cumulative sum of absolute forecast errors looks very similar for all three approaches, for the long horizon the cumulative sum of absolute forecast errors is smaller for ALM than for conventional DMA or DMS for the entire sample. Furthermore, ALM exhibits fewer and smaller jumps than conventional DMA or DMS.

4. Conclusion

DMA has been used extensively for the purpose of economic forecasting as it addresses dynamically both model and parameter uncertainty. This paper extends this framework by considering ALM. ALM dynamically averages not only over the forecasts of the individual models but, in addition, dynamically averages over the time-varying parameters. Therefore it exploits the information of all models in the estimation of the time-varying parameters. This is done by approximating the mixture of individual posteriors with a single posterior, which is then used in the upcoming period as the prior for each of the individual models. The relevance of this extension is illustrated in three empirical applications involving US inflation, US consumption expenditures and five major US exchange rate returns. It turns out that ALM leads to improved out-of-sample predictions in all applications.

References

Aye, G., Gupta, R., Hammoudeh, S., and Joong, W. (2015). Forecasting the price of gold using dynamic model averaging. *International Review of Financial Analysis*, 41:257–266.

Baur, D., Beckmann, J., and Czudaj, R. (2016). A metling pot - Gold price forecasts under model and parameter uncertainty. *International Review of Financial Analysis*, 48:282–291.

Beckmann, J. and Schüssler, R. (2016). Forecasting exchange rates under parameter and model uncertainty. *International Money and Finance*, 60:267–288.

Bork, L. and Moller, S. (2015). Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. *International Journal of Forecasting*, 31(1):63–78.

Bruyn, R., Gupta, R., and Eyden, R. (2015). Can we beat the randomwalk model for the south african rand-u.s. dollar and south african rand-uk pound exchange rates? evidence from dynamic model averaging? *Emerging Markets Finance and Trade*, 51(3):502–524.

Buncic, D. and Moretto, C. (2015). Forecasting copper prices with dynamic averaging and selection models. *The North American Journal of Economics and Finance*, 33:1–38.

Byrne, J., Korobilis, D., and Ribeiro, P. (2018). On the sources of uncertainty in exchange rate predictability. *International Economic Review*, 59(1):1468–2354.

Diebold, F. X. and Marion, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Drachal, K. (2016). Forecasting spot oil price in a dynamic model averaging framework. Have the determinants changed over time? *Energy Economics*, 60:35–46.

Fagin, S. (1964). Recursive linear regression theory, optimal filter theory, and error analyses of optimal systems. *IEEE International Convention Record Part 1*, pages 216–40.

Filippo, G. D. (2015). Dynamic model averaging and cpi inflation forecasts: A comparison between the euro area and the united states. *Journl of Forecasting*, 34(8):619–648.

Gupta, R., Hammoudeh, S., Kim, W., and Simo-Kengne, B. (2014). Forecasting china's foreign exchange reserves using dynamic model averaging: The roles of macroeconomic fundamentals, financial stress and economic uncertainty. *The North American Journal of Economics and Finance*, 28:170–189.

Hodrick, E. and Prescott, E. (1996). Postwar us business sycles: an empirical investigation. *Journal of Money, Credit and Banking*, pages 1–6.

Jazwinsky, A. (1970). *Stochastic Processes and Filtering Theory.* New York: Academic Press.

Koop, G. and Korobilis, D. (2011). Uk macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling*, 28(5):2307–2318.

Koop, G. and Korobilis, D. (2012). Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review*, 53(3):867–886.

Lauritzen, S. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420):1098–1108.

Liu, J., Wei, Y., Ma, F., and Wahab, M. (2017). Forecasting the realized range-based volatility using dynamic model averaging approach. *Economic Modelling*, 61:12–26.

Liu, L., Ma, F., and Wang, Y. (2015). Forecasting excess stock returns with crude oil market data. *Energy Economics*, 48:316–324.

Molodtsova, T. and Papell, D. (2009). Out-of-sample exchange rate predicitability with Taylor rule fundamentals. *Journal of International Economics*, 77(2):167–180.

Naser, H. (2016). Estimating and forecasting the real prices of crude oil: A data rich model using dynamic model averaging (dma) approach. *Energy Economics*, 56:75–87.

Raftery, A., Kárný, M., and Ettler, P. (2010). Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics*, 52(1):52–66.

RiskMetrics (1996). Technical document-fourth edition. Technical report, Available at: http://www.riskmetrics.com/system/files/private/td4e.pdf.

Risse, M. and Kern, M. (2016). Forecasting house-price growth in the euro area with dynamic model averaging. *The North American Journal of Economics and Finance*, 38:70–85.

Taylor, J. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39:195–214.

Timmermann, A. (2006). Forecast combinations. In Elliot, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1. North-Holland, Amsterdam.

Wang, Y., Ma, F., Wei, Y., and Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking and Finance*, 64:136–149.

Wei, Y. and Cao, Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from china. *Economic Modelling*, 61:147–155.

Appendix A. Tables

Table A.1: Forecasting performance for one quarter and one year inflation

		$h = 1$		$h = 4$	
Model	Hyerparamter	RMSFE	MAFE	RMSFE	MAFE
DMA	grid	0.35	0.21	1.10***	0.84
DMS	grid	0.40	0.24	1.09***	0.86
ALM	grid	0.36	0.23	1.00	0.80
DMA	0.99	0.37	0.22	1.22***	0.92
DMS	0.99	0.36	0.23	1.27***	0.96
ALM	0.99	0.35	0.22	1.05	0.82
DMA	0.95	0.36	0.23	1.08***	0.83
DMS	0.95	0.40	0.25	1.22***	0.97
ALM	0.95	0.37	0.23	1.00	0.80

The table shows the RMSFE and MAFE in percentage points for three different settings of the forgetting factor α (controls the change between models) and λ (controls the change in coefficients). In the first ($\alpha = \lambda = 0.95$), in the second ($\alpha = \lambda = 0.99$) and in the third both are estimated from a small grid $\alpha, \lambda \in (0.95, 0.99)$. The DM test calculates the statistic for the null hypotheses of equal squared forecast errors between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

Table A.2: Forecasting performance for one quarter and one year consumption expenditures

		$h = 1$		$h = 4$	
Model	Variable	RMSFE	MAFE	RMSFE	MAFE
DMA	nominal	0.59	0.38	2.36***	1.54
DMS	nominal	0.59	0.38	2.75***	1.74
ALM	nominal	0.63	0.38	2.02	1.23
DMA	real	0.42	0.33	1.64***	1.25
DMS	real	0.44	0.34	1.68***	1.29
ALM	real	0.42	0.32	1.42	1.04

The table shows the RMSFE and MAFE in percentage points for nominal and real US consumption expenditures. The DM test calculates the statistic for the null hypotheses of equal squared forecast error between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

Table A.3: Forecasting performance for one month and one year exchange rate

		$h = 1$		$h = 12$	
Model	Country	RMSFE	MAFE	RMSFE	MAFE
DMA	Canada	2.12*	1.50	11.69***	7.59
DMS	Canada	2.24**	1.57	13.14***	8.26
ALM	Canada	2.04	1.45	7.13	5.37
DMA	Denmark	3.17*	2.40	21.21***	14.81
DMS	Denmark	3.35***	2.51	21.71***	15.65
ALM	Denmark	3.10	2.39	12.39	10.33
DMA	United Kindom	2.99	2.22	15.63***	12.35
DMS	United Kindom	3.00	2.24	16.30***	12.67
ALM	United Kindom	2.97	2.24	11.06	08.62
DMA	Japan	3.42*	2.57	20.93***	14.85
DMS	Japan	3.59	2.70	13.14***	16.11
ALM	Japan	3.28	2.49	12.39	10.03
DMA	Sweden	3.23	2.44	17.90***	13.35
DMS	Sweden	3.43***	2.54	19.98***	14.69
ALM	Sweden	3.21	2.41	13.48	10.82

The table shows the RMSFE and MAFE in percentage points for five major US exchange rates returns. The DM test calculates the statistic for the null hypotheses of equal squared forecast errors between conventional DMA/DMS and ALM. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

Appendix B. Figures

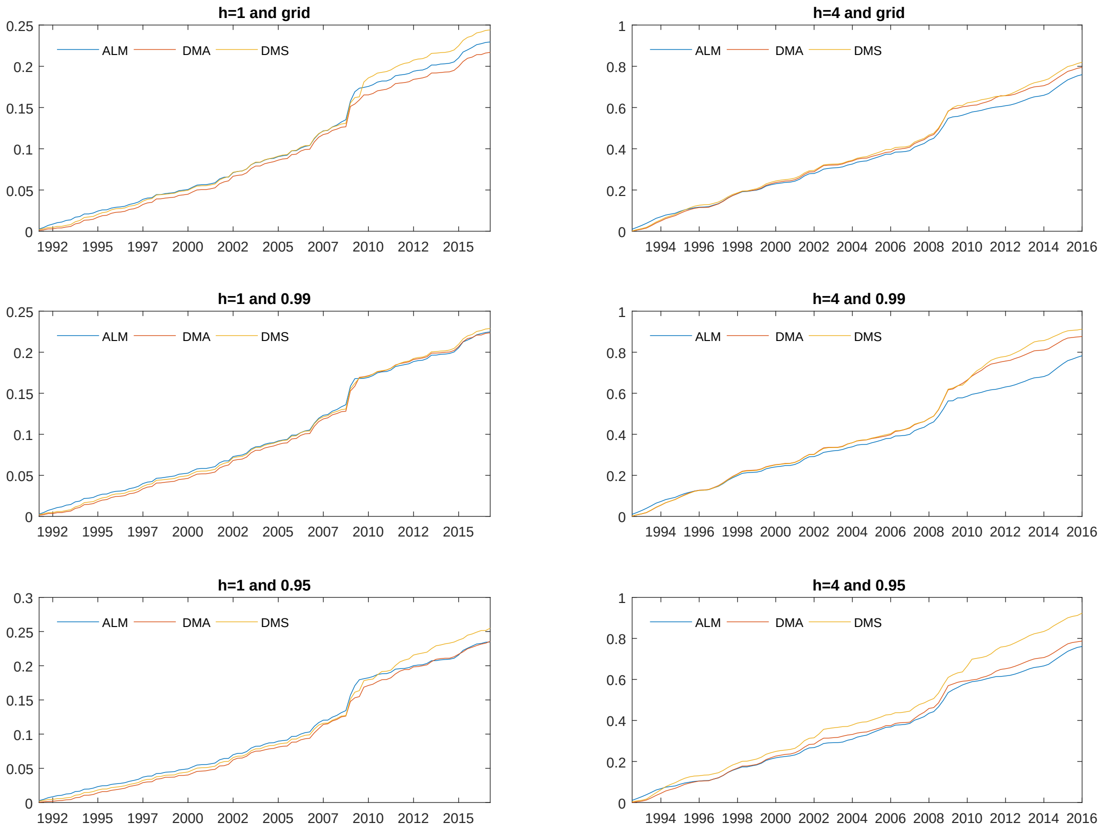

Figure B.1: Cumulative sum of absolute forecast errors for inflation.

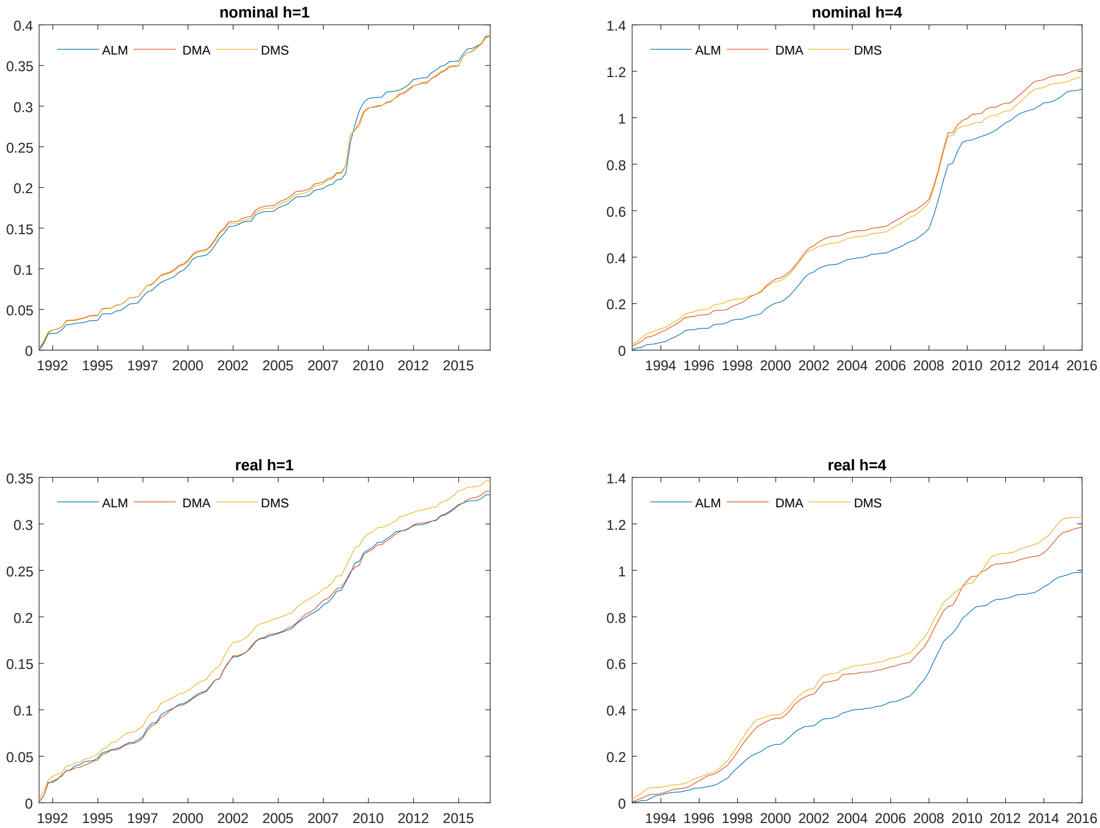

Figure B.2: Cumulative sum of absolute forecast errors for consumption expenditures.

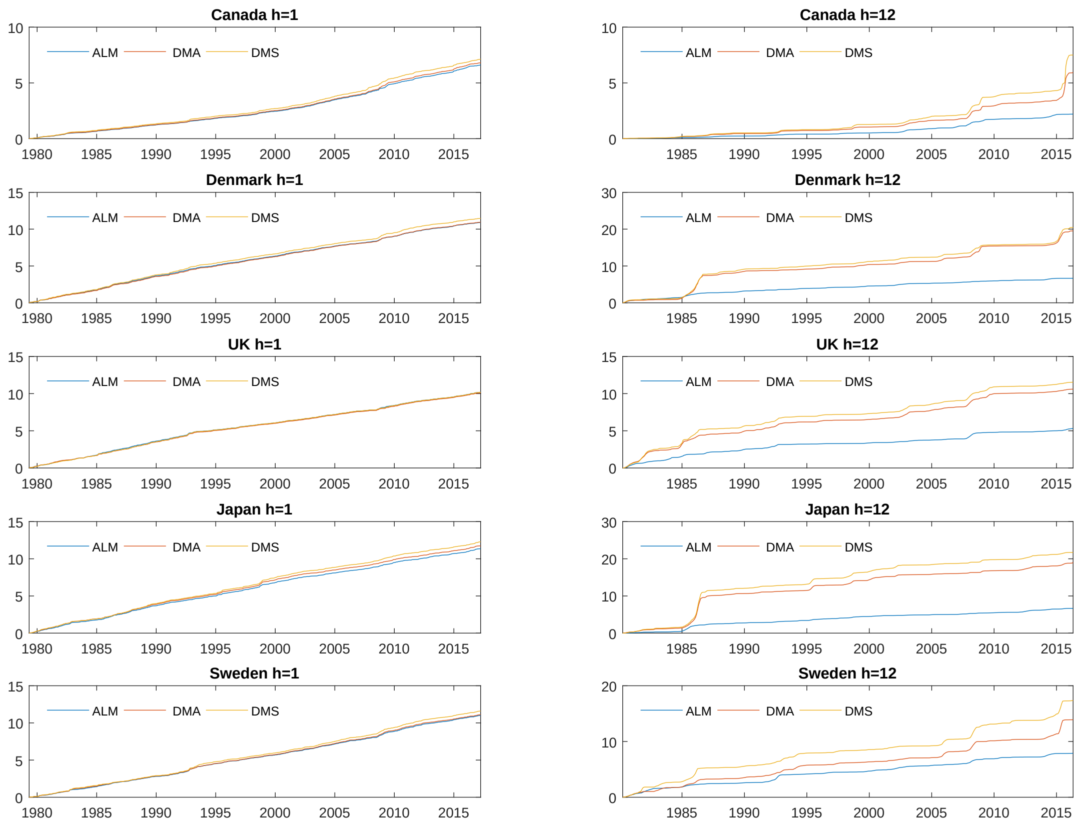

Figure B.3: Cumulative sum of absolute forecast errors for exchange rates.