# Adaptive Dynamic Model Averaging for House Price Forecasting

A. Yusupova and N.G. Pavlidis and E.G. Pavlidis

Lancaster University Management School

June 18 2019

# Introduction

Dynamic Model Averaging (DMA)

- Dynamic Linear Model: Time-variation in coefficients
- Adaptive model combination: Time-variation in model specification

DMA gained increased popularity in recent years for forecasting:

- inflation [Koop and Korobilis, 2012, Catania and Nonejad, 2018]
- carbon prices [Koop and Tole, 2013]
- exchange rates [Byrne et al., 2018]
- equity returns [Dangl and Halling, 2012]
- property price growth [Bork and Møller, 2015]

# Dynamic Linear Models with Forgetting

Dynamic Linear Model (Kalman Filter)

$$\begin{aligned} \theta_t^{(k)} &= \theta_{t-1}^{(k)} + \omega_t^{(k)}, \qquad & \omega_t^{(k)} &\sim \mathcal{N}\left(0, W_t^{(k)}\right) \\ y_t &= x_t^{(k)\top}\theta_t^{(k)} + \varepsilon_t^{(k)}, & \varepsilon_t^{(k)} &\sim \mathcal{N}\left(0, V_t^{(k)}\right) \end{aligned}$$

Estimation of $W_t^{(k)}$ poses computational problem (large $d$, large $K$)

**Forgetting factor** approach:

$$W_t^{(k)} = \frac{1-\lambda}{\lambda} C_{t-1}^{(k)}, \quad \lambda \in (0, 1]$$

Coefficient uncertainty associated with high coefficient variability

Choice of $\lambda$ **critically** affects forecasting performance

# Adaptive Forgetting DLM

**Central Idea**:

$$\lambda^\star = \underset{\lambda \in (0,1]}{\arg\min}\, \mathbb{E}_{Y,X}\left[\frac{1}{2}(y_{t+1} - \hat{y}_{t+1})^2\right]$$

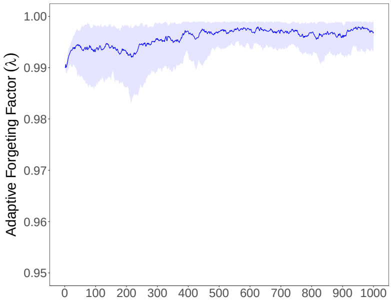
Function $\mathbb{E}_{Y,X}\left[\frac{1}{2}(y_{t+1} - \hat{y}_{t+1})^2\right]$ not observed (unfortunately)
Instead realisations we observe

$$J_{t+1} = \frac{1}{2}\left(y_{t+1} - \hat{y}_{t+1}\right)^2$$

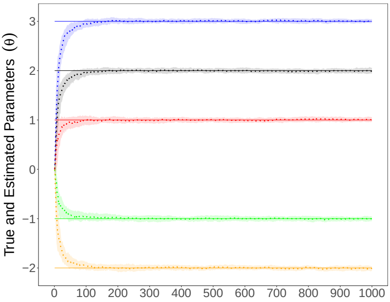
Stochastic optimisation: Use gradient $\frac{\partial J_{t+1}}{\partial \lambda}$ as approximation of gradient of expectation of squared forecast error

$$\frac{\partial J_{t+1}}{\partial \lambda} = \frac{\partial J_{\hat{\theta}_t}}{\partial \lambda}\frac{\partial \hat{\theta}_t}{\partial \lambda} = -(y_{t+1} - \hat{y}_{t+1})x_{t+1}^\top \frac{\partial \hat{\theta}_t}{\partial \lambda}$$

# Adaptive Forgetting DLM




(a) Median and IQR of stochastic gradient descent (SGD) estimates of $\lambda_t$




(b) Median and IQR of $\hat{\theta}_t$

Figure: Static Environment, $\theta_t = \theta_0$

# Adaptive Forgetting DLM

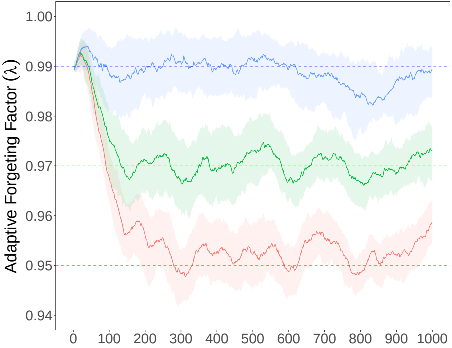
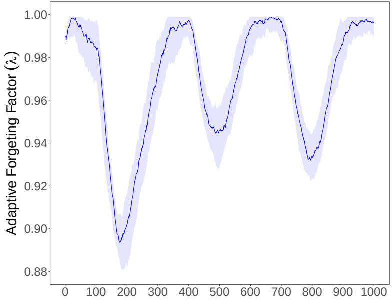
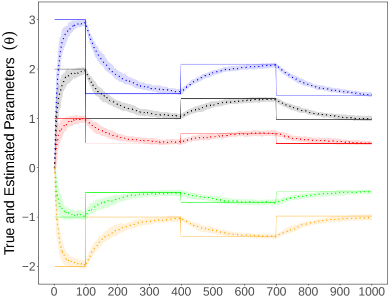



Figure: Median and IQR estimates of $\lambda_t$ through SGD for data simulated from a DLM with forgetting with $\lambda \in \{0.99, 0.97, 0.95\}$

# Adaptive Forgetting DLM




(a) Median and IQR of SGD estimates of $\lambda_t$




(b) Evolution of $\theta_t$ and median and IQR of $\hat{\theta}_t$

Figure: Coefficient vector subject to abrupt changes of different magnitute

No fixed value of $\lambda$ is appropriate for entire length of this time-series

# Model Combination

Inspired from Bayesian Model Average (BMA)

$$p(M_i|\mathcal{F}_t) \propto \underbrace{p(y_t|M_i, \mathcal{F}_{t-1})}_{\text{likelihood}} \left( \underbrace{p(M_i|\mathcal{F}_{t-1})^{\alpha}}_{(\text{prior of } M_i)^{\alpha}} + c \right)$$

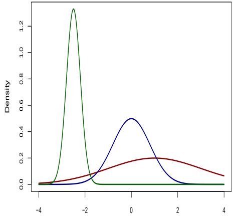
$$\propto p(M_i|\mathcal{F}_0)^{\alpha^t} \prod_{j=1}^{t} p(y_j|M_i, \mathcal{F}_{j-1})^{\alpha^{t-i}}, \quad \text{for } c = 0$$

$\alpha$: Forgetting factor for model combination stage

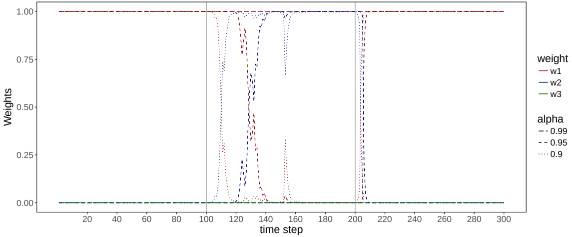
$\alpha$ is either 1 or very close to that (e.g. $\alpha = 0.99$)

Only Raftery et al. [2010] mention $c$

# Forgetting in Model Averaging can Fail



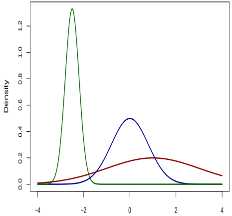

Normal densities



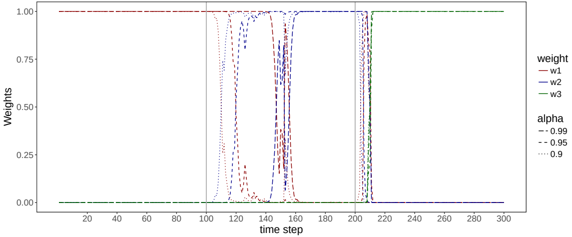

$c = 0$

$c = 0$ Fails to adapt to 2nd change point for all $\alpha$

# Forgetting in Model Averaging can Fail




Normal densities

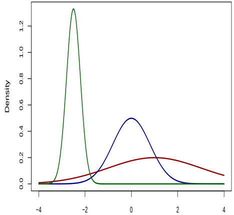



$c = 10^{-20}$

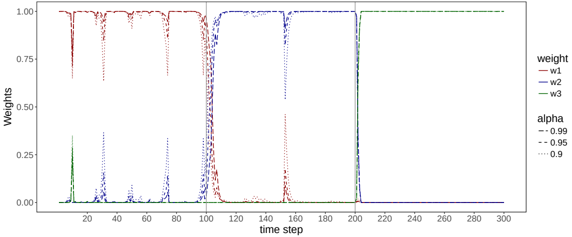
$c = 0$ Fails to adapt to 2nd change point for all $\alpha$

$c = 10^{-20}$ Speed of adaptation depends critically on $\alpha$

# Forgetting in Model Averaging can Fail




Normal densities




$c = 10^{-3}/K$

$c = 0$ Fails to adapt to 2nd change point for all $\alpha$

$c = 10^{-3}/K$ Adaptation for all $\alpha$ at cost of variability

$c > 0$ **at least as important as** $\alpha$

# Prediction with Expert Advice

## Online learning problem

Predict $y_{t+1}$ based on experts' forecasts $\hat{y}_{t+1} = \sum_{k=1}^{K} w_{t,k}\, \hat{y}_{t+1}^{(k)}$

**No assumptions** on time-series or forecasters

Assumptions on **loss function** $\ell$ (typically $\ell$ is bounded)

**Objective**: Design weight updates to minimise **regret** w.r.t. best sequence of $M$ experts (each optimal in a subset of the series)

$$R_T = \sum_{t=1}^{T} \ell(y_t, \hat{y}_t) - \sum_{m=1}^{M-1} \min_{k \in K} \left\{ \sum_{t=t(m)}^{t(m+1)-1} \ell(y_t, \hat{y}_t^{(k)}) \right\},$$

where $1 = t(1) < \cdots < t(M-1) < t(M) = T$ and $M$ are **unknown**

### ConfHedge

(2017) First algorithm with bound on $R_T$ for finite $T$ and **unbounded** $\ell$

ConfHedge algoritm involves **no** parameters

# Empirical Results (I)

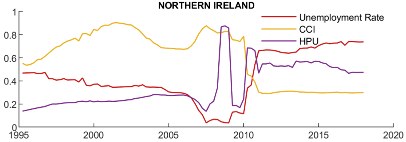
| Region | AR(1) | ADMA | eDMA | $DMA^{0.99}$ | $DMA^{0.95}$ | BMA | TVP |
|---|---|---|---|---|---|---|---|
| East Anglia | 101.05 | **0.80** | 0.92 | 1.03 | 1.24 | 1.32 | 1.54 |
| East Midlands | 70.74 | 0.80 | **0.78** | 0.84 | 0.83 | 0.79 | 1.55 |
| Greater London | 117.29 | 0.74 | 0.78 | 0.79 | 0.77 | **0.73** | 0.76 |
| Northern Ireland | 292.44 | **0.89** | 0.99 | 0.92 | 1.00 | 1.02 | 1.06 |
| North | 167.39 | 0.72 | 0.71 | 0.78 | 0.74 | **0.69** | 0.89 |
| North West | 64.59 | **0.93** | 0.94 | 1.02 | 0.94 | 0.94 | 1.13 |
| Outer Metropolitan | 54.51 | **0.95** | 0.96 | 1.14 | 0.96 | **0.95** | 1.27 |
| Outer South East | 61.84 | **0.92** | **0.92** | 1.06 | 1.05 | 1.06 | 1.42 |
| Scotland | 78.18 | 0.94 | **0.88** | 0.99 | 1.03 | 0.96 | 1.05 |
| South West | 60.25 | 0.99 | **0.91** | 0.98 | 1.15 | 1.13 | 1.75 |
| West Midlands | 53.34 | **0.85** | 0.90 | 0.95 | 0.94 | 0.94 | 1.55 |
| Wales | 148.29 | 0.81 | **0.79** | 0.90 | 0.80 | 0.80 | 0.87 |
| Yorkshire & Humber | 91.75 | **0.78** | 0.79 | 0.88 | 0.80 | 0.79 | 0.82 |

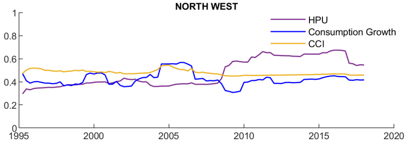
# Empirical Results (II)

| Region | eDMA | DMA$^{0.99}$ | DMA$^{0.95}$ | BMA | TVP |
|---|---|---|---|---|---|
| East Anglia | **0.88** | **0.79** | **0.66** | **0.61** | **0.52** |
| East Midlands | 1.03 | **0.96** | **0.97** | 1.02 | **0.52** |
| Greater London | **0.96** | **0.93** | **0.96** | 1.01 | **0.98** |
| Northern Ireland | **0.90** | **0.95** | **0.87** | **0.88** | **0.84** |
| North | 1.02 | **0.91** | **0.97** | 1.03 | **0.81** |
| North West | **0.99** | **0.91** | **0.99** | **0.99** | **0.82** |
| Outer Metropolitan | **0.99** | **0.84** | **0.99** | 1.00 | **0.75** |
| Outer South East | **0.99** | **0.87** | **0.89** | **0.86** | **0.64** |
| Scotland | 1.06 | **0.95** | **0.91** | **0.98** | **0.89** |
| South West | 1.09 | 1.01 | **0.86** | **0.88** | **0.57** |
| West Midlands | **0.94** | **0.89** | **0.89** | **0.89** | **0.54** |
| Wales | 1.03 | **0.89** | 1.00 | 1.01 | **0.93** |
| Yorkshire & Humber | **0.99** | **0.89** | **0.98** | **0.99** | **0.96** |

Ratios of MSFEs of ADMA relative to other forecasting models

# Uncertainty Index




## Key Findings

- Variable importance differs across regions and changes over time
- House Price Uncertainty: Constructed by text analysis of articles of 5 major newspapers obtained from LexisNexis
  - One of very few variables found important across all regions
  - HPU Movements coincide with recessions, and events such as the Brexit referendum and the negotiations of the withdrawal aggreement

# Summary

- ADMA: stochastic approximation to adapt $\lambda$, and ConfHedge algorithm for setting model weights
- ADMA: more accurate forecasts than AR(1) and all competing DMA specications
- No single predictive variable stands out as being the most important
- House Price Uncertainty index plays important role on the eve of 2008:Q3 price collapse and ahead of the EU Referendum