OIL PRICE FORECASTING WITH SOME GENETIC ALGORITHM VARIABLE SELECTION MODEL

Krzysztof Drachal
Faculty of Economic Sciences,
University of Warsaw,
Poland

Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021.

Motivation

- Time-series forecasting (for instance, spot oil price)
- Numerous potentially important explanatory variables
- Important explanatory variables vary in time
- The impact of explanatory variable on the forecasted time-series can also vary in time
- Time-Varying Parameters (recursive estimations, continuous income of new data from markets)
- Model averaging techniques – forecast combination instead of selecting only one, the “best”, model

Sketch of DMA

n – the number of potentially important explanatory variables; **$K = 2^n$ different linear regression models** can be constructed (including the model only with the intercept term)

y_t – the forecasted time-series; $x_t^{(k)}$ – the vector of explanatory variables in k-th linear regression model; k = { 1, ... ,K }

The state space model is given by:

$$y_t = x_{t-1}^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)},$$

$$\theta_t^{(k)} = \theta_{t-1}^{(k)} + \delta_t^{(k)},$$

with $\theta_t^{(k)}$ - the regression coefficients in the k-th linear regression model. Errors are normally distributed, i.e., $\varepsilon_t^{(k)} \sim N(0,V_t^{(k)})$ and $\delta_t^{(k)} \sim N(0,W_t^{(k)})$.

The DMA (Dynamic Model Averaging) forecast $\widehat{y_t^{DMA}}$ is formulated as $\boldsymbol{\widehat{y_t^{DMA}} = \sum_{k=1}^{K} \pi_{t|t-1,k} \widehat{y_t^{(k)}}}$, where $\widehat{y_t^{(k)}}$ is the forecast produced by the k-th regression model.

The **set of two weights is recursively updated** in the following way:

$$\pi_{t|t-1,k} = {(\pi_{t-1|t-1,k}^{\alpha})} \Big/ {(\sum_{i=1}^{K} \pi_{t-1|t-1,i}^{\alpha})},$$

$$\pi_{t|t,k} = {[\pi_{t|t-1,k} f_k(y_t | Y^{t-1})]} \Big/ {[\sum_{i=1}^{K} \pi_{t|t-1,i} f_i(y_t | Y^{t-1})]},$$

where $f_k(y_t|Y_{t-1})$ is the predictive density of the k-th model at y_t, and α is a so-called forgetting factor. The updating requires also another forgetting factor λ. Herein, α = 0.99 = λ is taken. $\pi_{0|0,k}$ is set to 1 / K. $\theta_t^{(k)}$ are updated with the Kalman filter and $V_0^{(k)} = 1$ and $W_0^{(k)} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & 1 & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$ are taken.

Problem: K(n) grows exponentially.

Sketch of GA-DMA

<u>Step 0:</u> Set up **the maximum number of component models** N.pop, mutation probability p.mut and crossover probability p.cross.

<u>Step 1:</u> Construct the initial population of component models mods.incl(t_0) as

- Random N.pop models out of all K possible ones, or
- Randomly generate by including or not including a given explanatory variable in N.pop models, or
- All possible one-variable models

Let there be n possible explanatory variables. Then individual component models are represented by 0-1 vectors of length n, in which each slot represents presence or absence of the given variable as an explanatory one.

<u>Step t:</u> Perform DMA over mods.incl(t) for data 1,…, t-1, and obtain weights $\pi_{t|t-1,k}$. Construct RVIs (Relative Variable Importances) for all explanatory variables, i.e., sums of $\pi_{t|t-1,k}$ from exactly those component models which contain the given variable as an explanatory one.

Construct mods.incl(t+1) by leaving in mods.incl(t) only models with N.pop highest weights $\boldsymbol{\pi_{t|t-1,k}}$.

Select in mods.incl(t+1) models to **mutate** (each model has the probability p.mut to mutate). In each of those models who will mutate randomly select exactly one explanatory variable slot. If this slot is 0, then with probability RVI corresponding to this variable, change this slot to 1.

Select models to **crossover** in mods.incl(t+1). Each model has the probability p.cross to be taken to crossover. If there will be selected odd number of models for the crossover, then randomly exclude one model out of the selected ones. Randomly pair those models. For each crossover pair of models cut in a random slot the vectors representing the models and crossover those two models. Expand mods.incl(t+1) by the new models obtained by crossover. (If mutation and crossover produces already existing model, then only one copy is passed to the next (sub-)step.)

<u>Step t +1:</u> Step t with **t → t + 1** .

Sketch of GA-DMA

n – number of potentially important explanatory variables; $K = 2^n$ linear regression models can be constructed, which for large n can lead to overwhelming number of models

Example: Linear regression model $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_n x_n$ can be encoded as a vector

[1, 1, 0, 0, … , 0, 1]

Example of **mutation**: $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_n x_n$ → $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_4 x_4 + \theta_n x_n$

[1, 1, 0, 0, … , 0, 1] → [1, 1, 0, 1, … , 0, 1]

Example of **crossover**: $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ and $y = \theta_0 + \theta_1 x_1 + \theta_3 x_3 + \theta_5 x_5$ make

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_5 x_5$ and $y = \theta_0 + \theta_1 x_1 + \theta_3 x_3$

[1, 1, 1, | 0, 0, 0 , … , 0] [1, 0, 1, | 0, 1 , 0 , … , 0]

[1, 1, 1, | 0, 1, 0 , … , 0] [1, 0, 1, | 0, 0 , 0 , … , 0]

Benchmarks

- TVP (Time-Varying Parameters linear regression with all potentially important explanatory variables, estimated similarly like each of the component models in DMA)
- NAIVE (no-change)

Variables

WTI – Spot price in USD per barrel

MSCI – MSCI World Index

TB3MS – U.S. 3-month treasury bill secondary market rate in %

CSP – Crude steel production in thousand tonnes

TWEXM – Trade weighted U.S. dollar index (Mar, 1973 = 100)

PROD – U.S. product supplied for crude oil and petroleum products in thousands of barrels

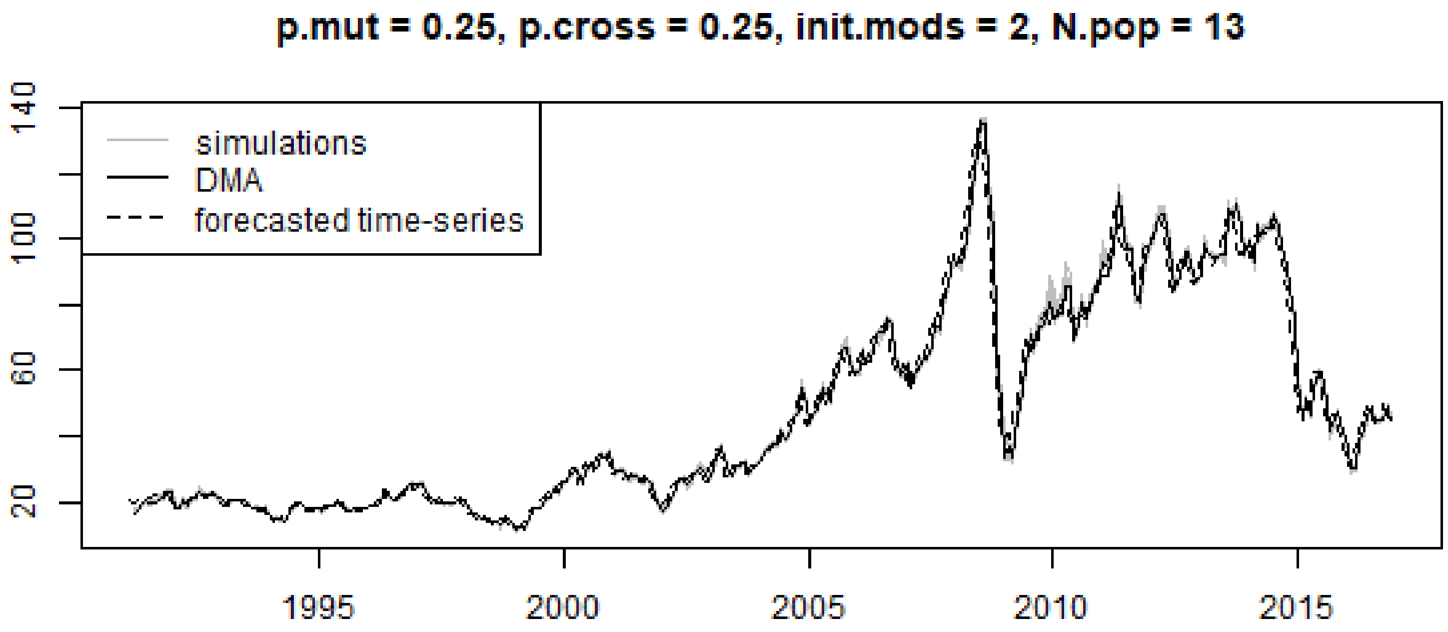
CONS – Total consumption of petroleum products in OECD in quad BTU

VXO – Implied volatility of S&P 100

Variables transformations and models parameters

- Monthly data between 02/1990 and 12/2016
- Explanatory variables lagged 1 period back
- Also log-diff(1 month) of WTI, MSCI and TWEXM, and log-diff(12 month) of CSP, PROD and CONS are taken
- Standardization based on mean and standard deviation based on in-sample period
- Observations 1, … , 100 are taken as in-sample (training) set
- Prob. of mutation: p.mut = {0.05, 0.25, 0.50, 0.75}
- Prob. of crossover: p.cross = {0.25, 0.50, 0.75, 0.90}
- Initial population: m = {0, 1, 2}
- Population size limit: N.pop = round(K*{0.10, 0.25}) = {13, 32}
- Number of simulations for each set of parameters: N.sim = 100

The selected example



Forecast accuracy measures

T – number of observations; y_t – real values; $\widehat{y_t}$ – forecasted values

- $\text{RMSE} = \sqrt{\sum_{t=1}^{T}(y_t-\widehat{y_t})^2 / T}$
- $\text{MAE} = \sum_{t=1}^{T}|y_t-\widehat{y_t}| / T$
- $\text{MASE} = \sum_{t=1}^{T}|y_t-\widehat{y_t}| \Big/ \left(\frac{T}{T-1}\sum_{t=2}^{T}|y_t-y_{t-1}|\right) = \text{MAE/ME(NAIVE)}$

DMA vs GA-DMA forecasts ($\varepsilon_t = \widehat{y_t^{DMA}} - \widehat{y_t^{GA-DMA}}$)

	mean	sd	variance	median	min	max	skew	kurtosis	coeff. of variation
RMSE	0.65	0.29	0.09	0.74	0.13	2.58	1.18	5.55	0.44
MAE	0.46	0.20	0.04	0.52	0.08	1.85	0.61	3.03	0.43
forecasted time-series (WTI)									
WTI	46.72	30.54	932.4	32.95	11.35	133.9	0.78	-0.68	0.65

Forecast accuracy measures

	mean	sd	variance	median	min	max	skew	kurtosis	coeff. of variation
GA-DMA (simulations with all parameters)									
RMSE	5.45	0.05	0.00	5.44	5.32	5.77	2.48	8.31	0.01
MAE	4.01	0.04	0.00	4.02	3.89	4.21	1.17	4.72	0.01
MASE	1.02	0.01	0.00	1.02	0.99	1.07	1.17	4.72	0.01

	mean	sd	variance	median	min	max	skew	kurtosis	coeff. of variation
forecasted time-series (WTI)	46.72	30.54	932.40	32.95	11.35	133.90	0.78	-0.68	0.65

	RMSE	MAE	MASE
DMA	5.41	3.96	1.01
TVP	5.44	4.03	1.03
NAIVE	5.53	3.93	1.00

Literature

- Drachal, K., 2020. Dynamic Model Averaging in economics and finance with fDMA: A package for R, Signals 1, 47-99
- Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy, International Journal of Forecasting 22, 679-688
- Koza, J. R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge: MIT Press
- Nonejad, N., 2021. An overview of Dynamic Model Averaging techniques in time-series econometrics, Journal of Economic Surveys 35, 566-614
- Onorante, L., Raftery, A. E., 2016. Dynamic model averaging in large model spaces using dynamic Occam's window, European Economic Review 81, 2-14
- Raftery, A. E., Karny, M., Ettler, P., 2010. Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill, Technometrics 52, 52-66
- Ramírez-Hassan, A., 2020. Dynamic variable selection in dynamic logistic regression: An application to Internet subscription, Empirical Economics 59, 909-932

Thank you for your attention!

kdrachal@wne.uw.edu.pl