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

- 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; \( k = \{1, \ldots, K\} \)

The state space model is given by:

\[
\begin{align*}
y_t &= x_{t-1}^{(k)} \theta_t^{(k)} + \epsilon_t^{(k)}, \\
\theta_t^{(k)} &= \theta_{t-1}^{(k)} + \delta_t^{(k)},
\end{align*}
\]

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

The DMA (Dynamic Model Averaging) forecast \( \hat{y}_t^{DMA} \) is formulated as

\[
\hat{y}_t^{DMA} = \sum_{k=1}^{K} \pi_t^{(k)} y_t^{(k)},
\]

where \( y_t^{(k)} \) is the forecast produced by the \( k \)-th regression model.

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

\[
\begin{align*}
\pi_{t|t-1,k} &= \frac{\alpha \pi_{t-1|t-1,k}}{\sum_{i=1}^{K} \pi_{t-1|t-1,i}}, \\
\pi_{t|t,k} &= \frac{\pi_{t|t-1,k} f_k(y_t | y_{t-1})}{\sum_{i=1}^{K} \pi_{t|t-1,i} f_i(y_t | y_{t-1})},
\end{align*}
\]

where \( f_k(y_t | y_{t-1}) \) is the predictive density of the \( k \)-th model at \( y_t \), and \( \alpha \) is a so-called forgetting factor. The updating requires also another forgetting factor \( \lambda \). Herein, \( \alpha = 0.99 = \lambda \) is taken. \( \pi_{0|0,k} \) is set to \( 1 / K \). \( \theta_t^{(k)} \) are updated with the Kalman filter and \( \Sigma_0^{(k)} = 1 \) and \( \Omega_0^{(k)} = \begin{bmatrix} 1 & \ldots & 0 \\ \vdots & 1 & \vdots \\ 0 & \ldots & 1 \end{bmatrix} \) are taken.

Problem: \( K(n) \) grows exponentially.
Step 0: Set up the maximum number of component models $N.\text{pop}$, mutation probability $p.\text{mut}$ and crossover probability $p.\text{cross}$.

Step 1: Construct the initial population of component models $\text{mods.incl}(t_0)$ as

- Random $N.\text{pop}$ models out of all $K$ possible ones, or
- Randomly generate by including or not including a given explanatory variable in $N.\text{pop}$ 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.

Step $t$: Perform DMA over $\text{mods.incl}(t)$ for data $1,\ldots, t-1$, 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 $\text{mods.incl}(t+1)$ by leaving in $\text{mods.incl}(t)$ only models with $N.\text{pop}$ highest weights $\pi_{t|t-1,k}$.

Select in $\text{mods.incl}(t+1)$ models to mutate (each model has the probability $p.\text{mut}$ 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 $\text{mods.incl}(t+1)$. Each model has the probability $p.\text{cross}$ 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 $\text{mods.incl}(t+1)$ 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.)

Step $t+1$: Step $t$ with $t \rightarrow t + 1$. 

Sketch of GA-DMA
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

\[ [1, 1, 0, 0, \ldots, 0, 1] \]

Example of **mutation**: \( y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_n x_n \)

\[ [1, 1, 0, 0, \ldots, 0, 1] \rightarrow [1, 1, 0, 1, \ldots, 0, 1] \]

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

\[ [1, 1, 1, | 0, 0, 0, \ldots, 0] \quad [1, 0, 1, | 0, 1, 0, \ldots, 0] \]

\[ [1, 1, 1, | 0, 1, 0, \ldots, 0] \quad [1, 0, 1, | 0, 0, 0, \ldots, 0] \]
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: \( p_{\text{mut}} = \{0.05, 0.25, 0.50, 0.75\} \)
- Prob. of crossover: \( p_{\text{cross}} = \{0.25, 0.50, 0.75, 0.90\} \)
- Initial population: \( m = \{0, 1, 2\} \)
- Population size limit: \( N_{\text{pop}} = \text{round}(K*\{0.10, 0.25\}) = \{13, 32\} \)
- Number of simulations for each set of parameters: \( N_{\text{sim}} = 100 \)
The selected example
Forecast accuracy measures

T – number of observations; \( y_t \) – real values; \( \hat{y}_t \) – forecasted values

- **RMSE** = \( \sqrt{\frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{T}} \)
- **MAE** = \( \frac{\sum_{t=1}^{T} |y_t - \hat{y}_t|}{T} \)
- **MASE** = \( \frac{\sum_{t=1}^{T} |y_t - \hat{y}_t|}{\frac{T}{T-1} \sum_{t=2}^{T} |y_t - y_{t-1}|} = \frac{\text{MAE}}{\text{ME(NAIVE)}} \)
DMA vs GA-DMA forecasts \( (\varepsilon_t = y_t^{DMA} - y_t^{GA-DMA}) \)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>variance</th>
<th>median</th>
<th>min</th>
<th>max</th>
<th>skew</th>
<th>kurtosis</th>
<th>coeff. of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.65</td>
<td>0.29</td>
<td>0.09</td>
<td>0.74</td>
<td>0.13</td>
<td>2.58</td>
<td>1.18</td>
<td>5.55</td>
<td>0.44</td>
</tr>
<tr>
<td>MAE</td>
<td>0.46</td>
<td>0.20</td>
<td>0.04</td>
<td>0.52</td>
<td>0.08</td>
<td>1.85</td>
<td>0.61</td>
<td>3.03</td>
<td>0.43</td>
</tr>
<tr>
<td>forecasted time-series (WTI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTI</td>
<td>46.72</td>
<td>30.54</td>
<td>932.4</td>
<td>32.95</td>
<td>11.35</td>
<td>133.9</td>
<td>0.78</td>
<td>-0.68</td>
<td>0.65</td>
</tr>
</tbody>
</table>
## Forecast accuracy measures

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

<table>
<thead>
<tr>
<th>forecasted time-series (WTI)</th>
<th>mean</th>
<th>sd</th>
<th>variance</th>
<th>median</th>
<th>min</th>
<th>max</th>
<th>skew</th>
<th>kurtosis</th>
<th>coeff. of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46.72</td>
<td>30.54</td>
<td>932.40</td>
<td>32.95</td>
<td>11.35</td>
<td>133.90</td>
<td>0.78</td>
<td>-0.68</td>
<td>0.65</td>
</tr>
</tbody>
</table>

| DMA | 5.41 | 3.96 | 1.01 |
| TVP | 5.44 | 4.03 | 1.03 |
| NAIVE | 5.53 | 3.93 | 1.00 |
Literature


• 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

Thank you for your attention!

kdrachal@wne.uw.edu.pl