An Automatic Identification Cycle for Dynamic Demand Model Generation

Alexander Efremov

Abstract: *In this paper a system identification cycle (SIC) for generation of the dynamic part of a price model is described. The reasons for modelling the dynamic sales behaviour are considered. The main problems with applying a standard identification procedure to real datasets are discussed. Then the SIC structure is presented. The applied data pre-processing techniques are mentioned and are considered the models and corresponded non-recursive methods estimating the dynamic price model. The validation criterion in SIC is defined and the results are presented. In the last section the results are analysed and according to the conclusions, the future work is outlined.*

Key words: *Dynamic Demand Model, Sales Forecasts, Identification Cycle.*

INTRODUCTION

The demand models are used to forecast unit sales. Based on the estimated sales, the retailers take decisions about their future actions, such as promotions, adds and displays. In this paper is presented an automated approach for obtaining of dynamic demand models.

Usually the effect of a given promotion carries over beyond the time period covered by the promotion. For instance, a promotion of canned food prompts people to overstock. Then to compensate, they buy less, before their normal pattern is restored. The demand models considered here provide an improved representation of the unit sales behaviour, taking into account the sales dynamics. The models are built into two steps. First step is to find appropriate static models that are used for initial forecasting the unit sales. They handle the main promotion factors (such as adds, displays, discounts...) and the existing cross effects between different products. The second step is determining additional regression models, accounting the dynamics, caused by the promotions. The input of a given regression model is the difference between the promotional forecast, computed by the corresponded static model and the baseline sales for the same product, i.e. an initial estimation of "stepwise" effect caused by running a promotion. The dynamic model output is the unit sales forecast.

The attention in the paper is focused on the second step – obtaining adequate dynamic descriptions, accounting the problems in models determination using real datasets. The main problems are big number of required models, different data lengths for each product, existence of wrong data in the datasets and not enough exciting input processes.

Usually the general set of data contains information about thousands of products, that makes impossible applying manual actions in the identification. This is the reason SIC to be realized as an automated procedure.

AUTOMATIC SYSTEM IDENTIFICATION CYCLE

As an initial analysis for identification method choice and model order determination cannot be made manually, a set of methods is used and for each method is applied a set of models with different orders. The resulting models are evaluated on the base of a validation criterion and the best dynamic model for each product is determined. To make an adequate assessment of the model accuracy, the datasets are divided into two parts. The first part is used for model determination and the second part – for validation. The remaining problems are solved by applying different restrictions on the data and pre-

processing techniques, before running the identification (see Fig. 1).

Data pre-processing

The observation intervals for each product are different. There are cases, where datasets have too small length, what makes them inappropriate for identification. Such datasets are removed by the following restriction on the data length N

$$N > \frac{1}{m}\left[N^* + \max(na, nb, nc, nd, nf)\right], \quad (1)$$

where m is the percentage value of the whole data used for identification, N^* is the minimal number of data used for identification, skipping the data for the initial conditions, accounted by the third term. The polynomials dimensions are denoted with na, nb, nc, nd and nf. The remaining datasets are detrended and then scaled by their standard deviations. An additional shaving procedure for removing wrong data of the output signals is applied. This pre-processing technique removes the spikes in the datasets and replaces them with interpolated values, determined by the neighbouring elements to the wrong data [1]. The above restriction and the mentioned techniques provide datasets appropriate for identification.

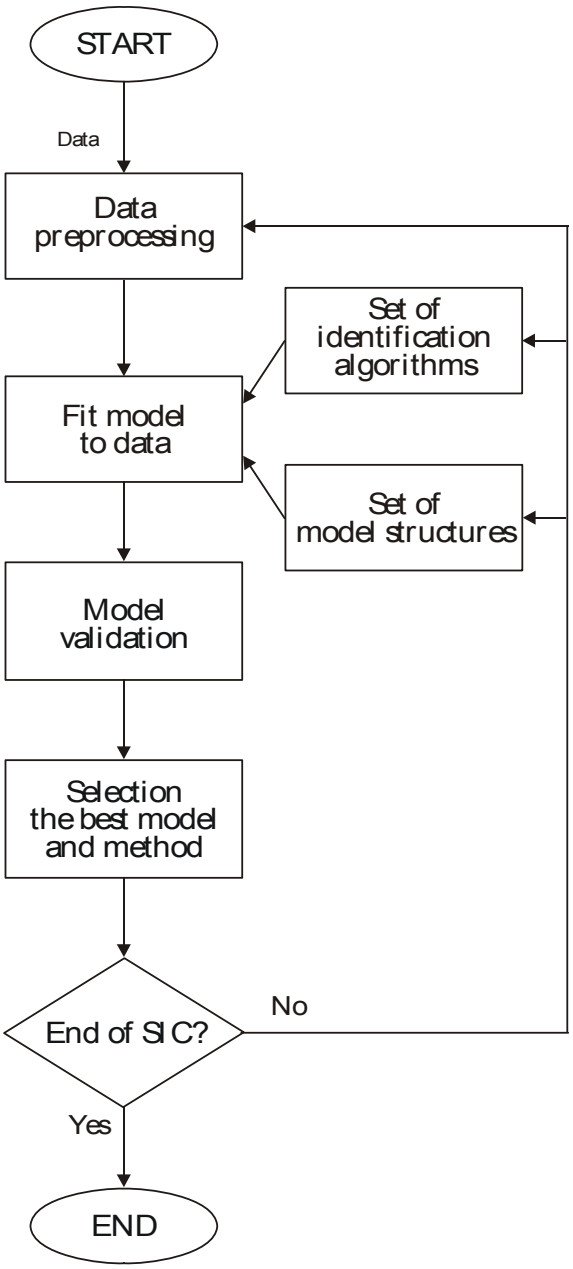

Fig 1. SIC structure

Identification Methods and Models

The set of non-recursive algorithms for system identification applied in SIC are [2]:

- Block Least Squares (BLS);
- Weighted Least Squares (WLS);
- Instrumental Variable Method (IVM);
- Correlated Least Squares (CLS);
- General Least Squares (GLS);
- Extended Matrix Least Squares (EMLS);
- Forecasting Error Method (FEM) applied for ARMAX model;
- FEM applied for OE model.

Block Least Squares

BLS, WLS, IVM and CLS are based on minimizing the least squares of the general error for the ARX model

$$A\left(q^{-1}\right)y_k = B\left(q^{-1}\right)q^{-d}u_k + e_k, \quad (2)$$

where y_k and u_k are the input and output in the k-th moment and $k \in [1; N]$. Generalizing the vector form of the above relation for the whole observation interval leads to the following matrix form

$$y = \Phi\theta + e, \tag{3}$$

where y is a vector including the output values for the whole observation interval, Φ is a matrix containing the input/output data and θ is a vector containing the polynomial parameters in (1), i.e.

$$\theta = [a_1 \quad a_2 \quad \dots \quad a_{na} \quad b_1 \quad b_2 \quad \dots \quad b_{nb}]^T,$$

The standard solution of the least square problem is

$$\hat{\theta} = \left(\Phi^T\Phi\right)^{-1}\Phi^T y.$$

For some datasets (if the input signal is not enough exciting), the matrix $\Phi^T\Phi$ is not well defined. To achieve more precise results, in SIC is applied a modification of the above relation. It is based on QR decomposition of $\Phi^T\Phi$. The modified solution is

$$\hat{\theta} = R^{-1}Q^T\Phi^T y,$$

Weighted Least Squares

The minimized cost function in BLS, which is a function of the general model error for the whole observation interval e, is

$$J_\theta = e^T e.$$

When the modelled dynamics is time variant it is appropriate to account the last information from the observation interval with greater weight with respect to the previous data. WLS is procedure that realizes this idea. The minimized criterion in WLS has the form

$$J_\theta = e^T W e,$$

where W is a weighted matrix usually chosen as a symmetric positive definite one. In the SIC are used six variants of W by which are defined different forgetting factors.

To achieve better results, the modification with QR decomposition, applied in BLS is also used. In this case QR is applied for the matrix $\Phi^T W\Phi$ and the parameters estimation becomes

$$\hat{\theta} = R^{-1}Q^T\Phi^T W y.$$

BLS and WLS provide non-biased parameter estimations if the process e is a white noise. Very often in practice the residual e has a significant colour part. One way to deal with the problem is to modify the generalized model (3). Another approach is to expand the model with a forming filter, which presents the colour error as a filtered white noise. The least squares problem in this case is applied for the corresponded extended model. The remaining algorithms are based on the above approaches and give a non-shifted solution for the case where the general error is a colour noise.

Instrumental Variable Method

IVM uses an additional instrumental matrix V. The elements of V have to be correlated with the elements of Φ, not correlated with the colour noise e and $\dim(V) \equiv \dim(\Phi)$. The matrix form of the model containing the whole data is

$$V^T y = V^T\Phi\theta + V^T e.$$

There are many variants for determination of an instrumental matrix. A four-step IVM is realized in SIC, where V depends on the past input data.

Correlated Least Squares

CLS is also applied when the general error is a colour noise. In this approach are used the estimation of cross correlation function between the processes u_k and y_k, denoted with R_{uy} and the estimation of auto correlation function of the input process, denoted with R_u.

The generalized model for that case is

$$R_{uy} = \Phi_R \theta + R_{ue}.$$

Φ_R contains the values of R_u and R_{uy} for different lags and R_{ue} is estimation of the cross correlation function between the processes u_k and e_k. To avoid a correlation between u_k and e_k, the minimal lag in the correlation functions is set to be greater then the lag of the auto-correlation function of the error R_e, for which R_e converges. The parameters estimate is

$$\hat{\theta} = R_R^{-1} Q_R^T \Phi_R^T R_{uy},$$

where $Q_R R_R = \Phi_R^T \Phi_R$.

General Least Squares

The conversion of the colour error $e_{c,k}$ into error, which has characteristics of white noise in GLS is made by the following forming filter

$$e_{c,k} = \frac{1}{D\left(q^{-1}\right)} e_k . \tag{4}$$

The extended model becomes ARARX

$$A\left(q^{-1}\right) y_k = B\left(q^{-1}\right) q^{-d} u_k + \frac{1}{D(q^{-1})} e_k . \tag{5}$$

GLS is realized as an iterative procedure, including two BLS methods, applied over two models. First model is the forming filter and the second one is the ARX model.

Extended Matrix Least Squares

The filter in EMLS has the structure

$$e_{c,k} = \frac{C\left(q^{-1}\right)}{D\left(q^{-1}\right)} e_k \tag{6}$$

The identified model in this case is ARARMAX

$$A\left(q^{-1}\right) y_k = B\left(q^{-1}\right) q^{-d} u_k + \frac{C(q^{-1})}{D(q^{-1})} e_k . \tag{7}$$

EMLS is also realized iteratively, but the two models (the input/output model and the filter) are identified simultaneously.

Forecasting Error Methods

All methods mentioned above are based on a linear approach for parametric model estimation. The last two FEM are non-linear approaches as the forecasting error is not linear with respect of θ. Here the process used in the criterion

$$J_\theta = e(\theta)^T e(\theta)$$

in moment k is

$$e_k(\theta) = y_k - \hat{y}_{k/k-1},$$

where $\hat{y}_{k/k-1}$ is the predicted output in moment k, calculated on the base of the past data by a model with parameters θ. As part of the regressors (past values of the forecasting error) also depends on θ, the process $\hat{y}_{k/k-1}$ is non-linear with respect of the model parameters. As the criterion cannot be minimized analytically an iterative procedure is used for minimizing of J_θ. A modification of the Gauss Newton method is applied, where the estimation of θ at the $i+1$-th iteration is

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \mu\left(J''_{\hat{\theta}^{(i)}}\right)^{-1} J'_{\hat{\theta}^{(i)}}.$$

$J'_{\hat{\theta}^{(i)}}$ is the gradient and $J''_{\hat{\theta}^{(i)}}$ is the Hessian of the cost function at the i-th iteration. Matrix μ is the step of the optimisation procedure, which is computing on each iteration [3].
Two FEM are used in SIC. One of them is deducted for ARMAX model

$$A\left(q^{-1}\right)y_k = B\left(q^{-1}\right)q^{-d}u_k + C(q^{-1})e_k. \quad (8)$$

and the other is based on OE model

$$y_k = \frac{B\left(q^{-1}\right)}{F\left(q^{-1}\right)}u_k + e_k. \quad (9)$$

Model Validation

As it was mentioned above, each dataset is divided into two parts – one for parameter determination and the other for model validation. The length of the first part is chosen to be 2/3 from the length of the observation interval, i.e. m in (1) is 66,7%. The remaining data is used for validation. The chosen validation criterion is variance accounted for (VAF) [1], which is a scaled version of the cost function J_θ:

$$VAF\left(y_k, \hat{y}_k\right) = \left(1 - \frac{\sum_{k=N^*+1}^{N}\left(y_k - \hat{y}_k\right)^2}{\sum_{k=N^*+1}^{N} y_k^2}\right)100\%.$$

When VAF is close to 100%, it corresponds to an accurate model. Otherwise the model doesn't fit well to data.

TEST DESCRIPTION AND TEST RESULTS

To investigate the developed software, a testing dataset containing real data for 50 products (weekly collected data for unit prices, seals, discounts, adds and displays for maximum 2 years) is used from the static models to generate the dynamic models inputs. For each product data are applied the mentioned above methods and for each method is applied a set of models with orders varying from 1 to 4. As the filter (4) in GLS requires few times greater order then the order of filter (6), the maximal dimension of $D\left(q^{-1}\right)$ for GLS is set to be 8. For each method and model structure the VAF criterion is used to assess the model accuracy. Finally, for every product is chosen the regression model with maximal VAF (MVAF). The values of MVAF are sorted in decreasing sequence and are shown on Fig. 2. For some datasets VAF has inadmissible values, regardless of the methods and applied model orders. Such case is shown on Fig. 3.

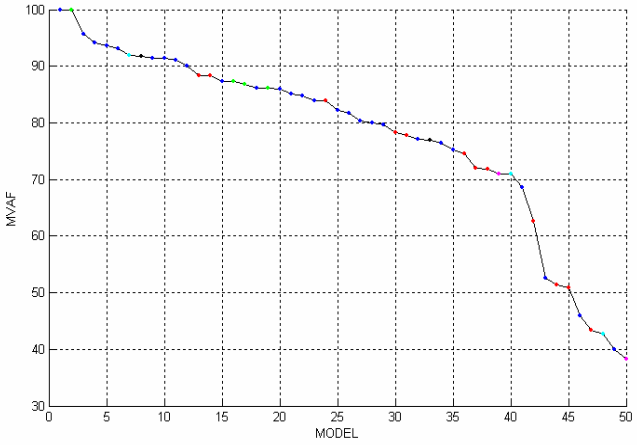

Fig 2 Sorted values of MVAF obtained for the dynamic models

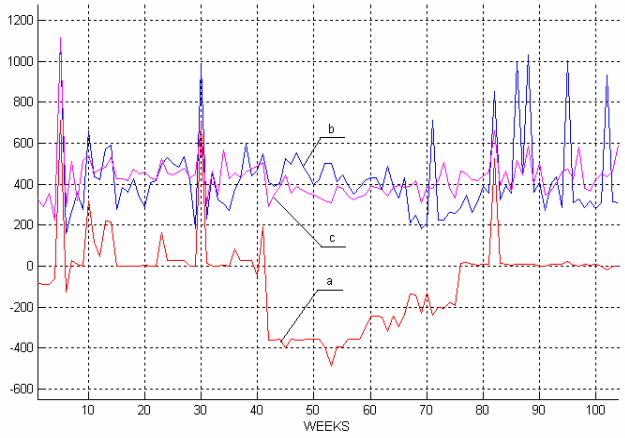

Fig 3. Input (a) and output (b) processes and model output (c), where MVAF = 51%

CONCLUSIONS AND FUTURE WORK

From the results shown above, the following conclusions and remarks are made:

In spite of the small data lengths and the bad data quality, in 80% of the cases the automated SIC provides appropriate models with VAF greater then 70%.

From Fig. 3 it is seen that there are areas, when the output behaviour is not correlated with the input process (see for example the spikes at the end of the output process). A reason for that is the influence of processes that are not added as input signals in the current models, such as cross effects between competitive products, units connected by a promotion or competitors actions. Part of the signals are available in the dataset and the model input should be extended with them. Other reason for the small values of VAF is short observation intervals for some datasets. Moreover, as the identification and validation uses subsets of the whole data, both the procedures in these cases are not precise. An improvement can be expected, if the whole observation interval is used for identification.

The future steps in improving the dynamic models are to apply recursive procedures for weekly updating the model parameters. As the real dynamics is usually non-stationary (because of competitors actions, adding new products, promotions...), different modifications will be made to provide enough sensitivity of the recursive algorithms to the current data.

To account the dynamic cross-effect between different products, the model input will be extended with the available significant actions on the unit sales.

REFERENCES

[1] Verhaegen, M., Verdult, V. Filtering and System Identification: An Introduction. TU–Delft, 2003.

[2] Гарипов, Е. Идентификация на системи. ТУ–София, 1997.

[3] Айдемирски, П. Изследване на методи и алгоритми за адаптивна обработка на сигнали. Дисертация. ТУ-София.

ABOUT THE AUTHOR

Ass. Prof. Alexander Efremov, Department of Computer Systems Applications, University of Chemical Technology and Metallurgy, Sofia, Phone: +359 02 8163-485, e-mail: alefremov@yahoo.com