BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **7**, No 2

Sofia • 2007

Automatic Multivariate System Identification Dealing with Problematic Datasets

Alexander Efremov, Assen Todorov

Technical University of Sofia, 1000 Sofia
E-mails: alefremov@tu-sofia.bg assent@tu-sofia.bg

Abstract: *An automatic system identification cycle (ASIC) for generation of multivariate dynamic models is designed. The reasons for modeling some multiple input multiple output (MIMO) systems without human intervention are discussed. The problems arising during the identification due to removing the manual actions and using unsuitable datasets are investigated. Appropriate solutions are suggested. The implemented preprocessing techniques, the set of model types and structures and the set of estimators are described. The automatic MIMO identification approach is used for modeling of market dynamics. To decrease the number of factors in the considered MIMO system, a procedure which determines a subset of only significant input-output relations is proposed. The MIMO models are compared with another dynamic approach, where the demand model contains a multivariate static and single input, single output (SISO) dynamic parts.*

Key words: *MIMO identification, market system, multivariate dynamic demand model.*

1. Introduction

Many real phenomena are related to each other and it is often hard or impossible to isolate particular signals relation from the whole set of interacting processes. On the other hand, taking into account the multivariate character of such kind of systems leads to more accurate representation of their behavior. In this paper an identification approach, based on a decomposition of the MIMO structure to a set of multiple input single output (MISO) subsystems is considered. The enormous number of inputs and outputs makes difficult the human intervention during the MISO models determination. To obtain a model, if there are hundreds of observed processes, it is necessary to avoid the manual actions and to have an automatic models generation.

The augmentation of the system inputs and outputs leads to increasing of the computational burden of the identification procedure. Also, as the number of models parameters grows, the subsequent procedures employing the model become more complicated. And finally, accounting for inputs that have no significant effect on the system behavior, may deteriorate the models accuracy. In order to decrease the influence of such problems, it is reasonable to determine the number of significant input-output relations before running the estimation procedures and then to take into account the resulting subset of input signals in the MISO models.

In some applications apriori information is not available or it is not enough to make assumptions for the correct models types and structures. A proper practical solution, when it is not possible to deduce initial mathematical descriptions, is to choose simple relations between the inputs and outputs. This is especially true for time-varying systems where an accurate model for a given time period is not necessary to be applicable for another period of time. In such case the convergence rate of the estimated parameters may not be enough to maintain a suitable representation. Because of these reasons we designed automatic system identification cycle (ASIC), which is based on linear regression models. The theory is well developed for these representations and it makes them easy to be applied in many practical cases. By a manipulation of the factors in the regression functions these models can also describe some types of non-linear behavior [10].

An additional requirement for ASIC is to have an open structure with respect to the sets of models types and structures and the set of estimators. If any apriori information is available, it may be useful to specify these sets in advance.

Sometimes the accomplishment of real experiments for the aim of system identification is connected with significant economic losses. Hence, it is important to take into account the available data for the system behavior and not to assign any additional experiments. Therefore ASIC should be able to manage with problematic observations, such as unrealistic values, missing data, short observation intervals or when the system is not excited enough. As the manual actions are removed from the identification process, it is necessary to design initial tests upon the datasets that reveal such problems. The irregular data has to be isolated and the datasets to be automatically repaired by appropriate preprocessing techniques in order to decrease the influence of such data records. When the problematic or missing data is replaced by posteriori determined values, an error remains, which in some cases becomes significant. To account for this kind of disparity and to provide a robust solution, suitable estimators modifications and their numerically stable realizations are used.

2. Data preprocessing

ASIC starts with an observation of the data size and quality. If the length of the observation interval is less than a certain limit, ASIC is stopped. This limit depends on the signal to noise ratio and the concrete model type and structure. Also the input processes are tested for satisfaction of the persistency of excitation condition. The improper inputs are removed from the consideration. After these checks, the following data preprocessing techniques are run:

- Filling empty records – introducing values in the dataset, if there is missing data;

•Resampling – if the sample time is not constant, the input-output data and the time instants are updated in such way that the time intervals between two observations are constant (this action is added to preserve the correct work of ASIC for the marketing system, described in section 4);

• Realistic values check – it consists of a shaving procedure, which determines the artifacts (spikes) in the datasets and replaces them by linearly interpolated values. Also, if apriori data is available, additional restrictions on the data can be defined.

•Detrending and deseasonalizing – a linear trend and seasonality (for the market identification) are identified preliminary and then removed from the available observations. Subsequently the obtained model from ASIC is combined with the trend and seasonality models.

•Scaling – the datasets are scaled by their standard deviations, so that the modified processes have unit variances.

The missing input-output observations may cause problems, which in some cases lead even to impossibility to accomplish the process of identification. In industry such cases are due to incorrect sensors work or problems with transferring the data. In market systems it is possible that a product is not on the market for a period of time; or the lack of data may be caused by database failures.

The MIMO identification requires complete information about the system behavior for the whole observation interval. If there are periods of time when parts of the observations do not exist, some values should be assigned. This incomplete information is a serious problem if the data length is short and experiments for the purposes of identification cannot be made. Thus it is important to take into account all the available information about the investigated behavior, but not to skip periods of time when the data is not full.

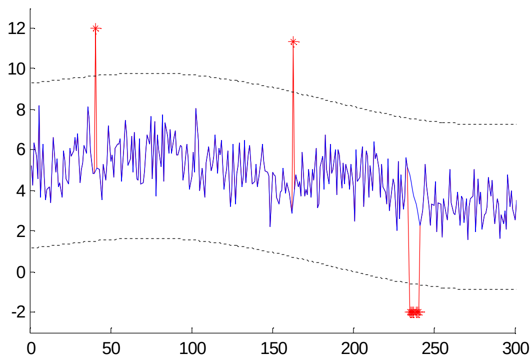

Fig. 1. Removing of unrealistic records

To discover the spikes in the data, we apply the following shaving procedure, based on [19]. At first a low-pass filter determines the trend $x_{t,k}$ of the investigated process x_k. After that, the standard deviation $\sigma_{x_{dt}}$ of the detrended signal $x_{dt,k} = x_k - x_{t,k}$ is computed. It determines the upper and lower limits, which distinguish the realistic values from the artifacts in the data. The range between these limits is defined as $n\sigma_{x_{dt}}$, where n is an initial parameter of ASIC, which is preliminary adjusted. If the processes distribution is known, n can be determined appropriately. For instance the confidential interval $x_{t,k} \pm 3\sigma_{x_{dt}}$ encloses 99.73% of the Gaussian distribution and the upper and lower limits of the process x_k could be chosen as

$$l_{\mathrm{upper},k} = x_{t,k} + 3\sigma_{x_{dt}},\; l_{\mathrm{lower},k} = x_{t,k} - 3\sigma_{x_{dt}}.$$

After the determination of these constraints, the process is checked for spikes. If incorrect values are found outside the limits, they are replaced by linear interpolated values (Fig. 1).

For instance, wrong records in the retail industry may be due to incorrect data aggregation.

The values added in the dataset only to provide appropriate work of the estimators are accounted with lower weight during the parameters estimation and cross effects determination.

3. MIMO model determination

There are different approaches to obtain MIMO representations. One of them is subspace identification [15, 17, 18], which is a powerful tool for modeling of multivariate systems. The subspace methods estimate state space models. They provide good initial estimation of the system behavior which does not depend on initial conditions. Usually to improve the estimates, these methods are combined with another procedure (based on minimization of the output error or the prediction error). Another way is based on estimation of regression models. The SISO system identification is considered in details in [10]. The presented ASIC is related to the second approach, since it is a continuation of [7], where we reduce the dynamics of a multivariate system to sets of MISO static and SISO dynamic models. A possible way to obtain MIMO representation of given dynamics, is to use one MIMO regression model, but in this case a problem connected with the choice of the model structure occurs. Also numerical problems can arise, which is a serious disadvantage, when the identification is applied automatically. To avoid this, it is possible to run l times the estimation procedure for MISO models that gives the relation between a given system output and all input processes. This approach avoids or decreases the influence of the problems above mentioned, which is the reason the MISO identification to be implemented in ASIC.

3.1. Set of estimators

The estimation procedures that we implemented in ASIC are multivariate versions of the following algorithms

- Least Squares (LS);
- Generalized Least Squares (GLS);
- Extended Matrix Least Squares (EMLS);
- Weighted Least Squares (WLS);
- Instrumental Variable (IV);
- Prediction Error (PE) applied for Auto-Regressive Moving Average models with eXogenous input (ARMAX) model;
- PE applied for Output Error (OE) model;
- PE applied for Box-Jenkins (BJ) model.

Only the LS estimator for MISO models determination is presented in details below. The other estimators are deducted in a similar way. The standard LS minimizes the sum of squared residual, which accounts for the presence of measurement noise, unfitness of the model to the real system behavior and the influence of the environment on the system dynamics. The model related to *j*-th output estimating by LS, is MISO-ARX (Auto-Regressive model with eXogenous input)

(1) $$A_j(q^{-1})y_{j,k} = B_{j1}(q^{-1})u_{1,k} + B_{j2}(q^{-1})u_{2,k} + \ldots + B_{jm}(q^{-1})u_{m,k} + e_{j,k},$$

where $u_{i,k}$ for $i=\overline{1,m}$ and $y_{j,k}$ for $j=\overline{1,l}$ are the system input and output respectively, $e_{j,k}$ is the residual process, q^{-1} is the unit time delay operator, $k \in [0; N-1]$, where N is the length of the observation interval and

$$A_j(q^{-1}) = 1 + a_{j,1}q^{-1} + \ldots + a_{j,na_j}q^{-na_j}$$

and

$$B_{ji}(q^{-1}) = b_{ji,1}q^{-1} + \ldots + b_{ji,nb_{ij}}q^{-nb_{ji}}$$

are the model polynomials. The parameter vector $\theta_j \in \mathrm{R}^{na_j + nb_{j1} + \ldots + nb_{jm}}$ is

$$\theta_j = [a_{j,1}\ a_{j,2}\ \ldots\ a_{j,na_j}\quad b_{j1,1}\ b_{j1,2}\ \ldots\ b_{j1,nb_{j1}}\quad b_{j2,1}\ b_{j2,2}\ \ldots\ b_{j2,nb_{j2}}\quad \ldots$$
$$b_{jm,1}\ b_{jm,2}\ \ldots\ b_{jm,nb_{jm}}]^{\mathrm{T}}.$$

The structure parameters na_j and nb_{ji} are the degrees of the model polynomials.

To simplify the explanation in this section, we will skip the subscript *j*, but the presented models will be related to the *j*-th output.

A generalization of the regression model for all observations leads to the following matrix form

(2) $$y = \Phi\theta + e,$$

where the column vectors y and e contain the last $N-n$ ($n = \max(na, nb)$) output observations and the corresponding residuals and $\Phi \in \mathrm{R}^{N-n\times\dim\theta}$ is the data matrix. The LS criterion related to the *j*-th output is

$$\min_\theta J_\theta = \min_\theta e^{\mathrm{T}}e, \text{ subject to (2).}$$

Actually J_θ is the 2-norm of the residual vector. So the LS criterion becomes

(3) $$\min_\theta J_\theta = \min_\theta \| \Phi\theta - y \|_2^2,$$

where the residual e is skipped from the consideration. The standard LS solution is

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

A generalization of LS is obtained by introduction of different weights for the residual values in the criterion. This is the idea of WLS. The criterion corresponding to a given MISO model is

$$\min_\theta J_\theta = \min_\theta \| y - \Phi\theta \|_W^2.$$

The matrix $W \in \mathrm{R}^{N-n\times(N-n)}$ is diagonal and contains the residual weights. The WLS solution is

$$\hat{\theta} = (\Phi^T W\Phi)^{-1}\Phi^T W y \cdot$$

The properties of WLS depend on the choice of the weighting matrix W. For instance, an optimization procedure, which minimizes the variance of e, when the residual is not white noise, is considered in [18]. Also WLS is suitable for time varying systems. Usually in such cases W is diagonal, with entries that decrease exponentially. WLS is included in the set of estimators where different variants of that correspond to concrete forgetting factors are realized.

If the residual is a color noise (in this case it will be denoted as e_{ck}), the resulting estimates are biased. One way to deal with this problem is to modify (2). Another way is to expand the model with a filter, which represents the residual as filtered white noise. The remaining algorithms are based on the above approaches and provide a non-shifted solution for the cases of color residuals.

The IV algorithm provides unbiased estimates by an instrumental matrix V. It is determined in such a way, that its elements are correlated with the regressors, but do not depend on the color noise e_{ck}. The relation between a given output and the system inputs in IV is modified as

$$V^{\mathrm{T}} y = V^{\mathrm{T}} \Phi\theta + V^{\mathrm{T}} e_c = V^{\mathrm{T}} \Phi\theta + e\,.$$

When the entries of the instrumental matrix are not correlated with e_c, the process $e = V^{\mathrm{T}} e_c$ has the character of a white noise. In ASIC, a four-step IV [10], estimating MISO-ARX models is realized, where V depends on the past input data.

The conversion of the color error e_c into a process, which has white noise characteristics in GLS, is made by the following SISO forming filter

$$e_{c,k} = D^{-1}(q^{-1}) e_k\,.$$

The resulting extended model, related to a given output is MISO-ARARX (ARX with Auto-Regressive noise model).

$$A(q^{-1}) y_k = B_1(q^{-1}) u_{1,k} + B_2(q^{-1}) u_{2,k} + \ldots + B_m(q^{-1}) u_{m,k} + D^{-1}(q^{-1}) e_k\,.$$

GLS is an iterative procedure, where LS is running two times at each iteration. First it is applied to the forming filter i.e. to estimate the parameters of $D(q^{-1})$ and after that LS is used to estimate $A(q^{-1})$ and $B(q^{-1})$ of the MISO-ARX model.

EMLS implemented in ASIC estimates MISO-ARARMAX (ARMAX with Auto-Regressive noise model). The filter in this case has the structure

$$e_{c,k} = D^{-1}(q^{-1}) C(q^{-1}) e_k$$

and the extended model is

$$A(q^{-1}) y_k = B_1(q^{-1}) u_{1,k} + B_2(q^{-1}) u_{2,k} + \ldots + B_m(q^{-1}) u_{m,k} + D^{-1}(q^{-1}) C(q^{-1}) e_k\,.$$

EMLS is also an iterative procedure, but in contrast to GLS, all model parameters are estimated simultaneously.

The above methods minimize the general error, which is a linear function of θ. The PE methods are non-linear as the prediction error $e_{p,k} = y_k - y_{k/(k-1)}$ is not linear in θ. The predicted output $y_{k/(k-1)}$ is calculated on the basis of the past data by a model with parameters θ. If unmeasured processes determined as functions of θ, take part in the regression vector, then $y_{k/(k-1)}$ and the prediction error become non-linear with respect to the estimated parameters. Let e_p is a vector containing the last $N - n$ values of the prediction error, then the cost function is

$$J_\theta = e_p^{\mathrm{T}} e_p ,$$

which cannot be minimized analytically. Therefore an iterative procedure should be used for J_θ minimization. In ASIC we applied Gauss-Newton method. The estimate of θ at $(h+1)$-st iteration is

$$\hat{\theta}^{h+1} = \hat{\theta}^{h} - \mu^h H_{\hat{\theta}^h}^{-1} J'_{\hat{\theta}^h} . \tag{5}$$

$J'_{\hat{\theta}^h}$ is the gradient of J_θ at the current iteration and $H_{\hat{\theta}^h}$ is an approximation of its Hessian. The cost function gradient and the Hessian approximation are

$$J'_\theta = \Psi_\theta e_p ,\ H_\theta = \Psi_\theta \Psi_\theta^{\mathrm{T}} ,$$

where $\Psi_\theta \in \mathrm{R}^{\dim\theta \times (N-n)}$ is a matrix with columns equal to the gradients of e_{pk} with respect to θ. The matrix μ^h is the step size of the optimization procedure, which is computed at each iteration [4] to improve the convergence speed and the precision of the iterative procedure.

Three PE methods are implemented. They are deduced for MISO-ARMAX model

$$A(q^{-1}) y_k = B_1(q^{-1}) u_{1,k} + B_2(q^{-1}) u_{2,k} + \ldots + B_m(q^{-1}) u_{m,k} + C(q^{-1}) e_k ,$$

MISO-OE model

$$y_k = F^{-1}(q^{-1}) (B_1(q^{-1}) u_{1,k} + B_2(q^{-1}) u_{2,k} + \ldots + B_m(q^{-1}) u_{m,k}) + e_k$$

and MISO-BJ model respectively

$$y_k = F^{-1}(q^{-1})(B_1(q^{-1}) u_{1,k} + B_2(q^{-1}) u_{2,k} + \ldots + B_m(q^{-1})) u_{m,k} + D^{-1}(q^{-1}) C(q^{-1}) e_k .$$

3.2. Estimators modifications

3.2.1. Numerically stable modifications

To solve the estimation problems in the previous version of ASIC [7], QR decomposition was used. Here, to obtain numerically reliable solution of the above methods we present an improved ASIC, in which according to [9, 18] the singular value decomposition (SVD) is applied to the data matrix Φ. The economy size SVD is

$$\Phi = U \Sigma V^{\mathrm{T}} ,$$

where Σ is square and diagonal. Taking into account the properties of the orthogonal matrices U and V, the LS solution becomes

$$\hat{\theta} = V \Sigma^{-1} U^{\mathrm{T}} y .$$

If Φ has linear dependent columns, the numerical problems can be avoided by using the above decomposition. Let $\mathrm{rank}(\Phi) = r$ and $r < \dim\theta$. Since Σ is a diagonal matrix that contains the singular values of σ_i arranged in a decreasing sequence, SVD applied to Φ can be presented as

$$\Phi = [U_1 \quad U_2] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix} = U_1 \Sigma_r V_1^{\mathrm{T}},$$

where the singular values in $\sum_r$ are greater than zero. So the modified LS solution becomes

$$\hat{\theta} = V_1 \Sigma_r^{-1} U_1^{\mathrm{T}} y .$$

If some columns of Φ are almost linear dependent, this implies that some singular values of $\sum_r$ are very small. To avoid this, a regularization of the above modification can be applied if the singular values σ_i are restricted from below by introducing a low limit $\sigma_{\min} > 0$. Thus r becomes equal to the number of singular values greater than or equal to $\sigma_{\min}$. The condition number κ of $\sum_r$ is $\kappa = \sigma_1 / \sigma_r$, so the restriction of the minimum singular value σ_r provides a well-conditioned matrix $\sum_r$. In accordance to the definition of κ, we compute $\sigma_{\min}$ as $\sigma_{\min} = \sigma_1 / \kappa_d$ with κ_d – a desired maximum condition number of $\sum_r$. The advantage of this regularization is that neglecting the smallest singular values, loosely speaking, we take into account only the dominating part of $\sum$ and skip the part that leads to ill-conditioned problem.

To achieve numerically stable solution of PE, it is necessary H_θ to be a positive definite matrix. It is possible the model to be insensitive to changes in some parameters and Ψ_θ to be ill-conditioned. An appropriate solution is to use the Levenberg-Marquardt regularization [14], where $H_\theta = \Psi_\theta \Psi_\theta^T + I\lambda$. Here $\lambda > 0$ is the regularization parameter, which decreases when the procedure converges. Such a modification is implemented in [7]. A disadvantage of the method is the bias, which appears for $\lambda \neq 0$. Another way to provide stable estimation is to modify (5) again by using SVD. According to the above considerations, if $\operatorname{rank}(\Psi_\theta) = r$ for $r < \dim \theta$ after applying SVD to Ψ_θ, we get

$$\Psi_\theta = [U_1 \quad U_2] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix}$$

and (5) becomes

$$\hat{\theta}^{h+1} = \hat{\theta}^h - \mu^h U_1 \Sigma_r^{-1} V_1^{\mathrm{T}} e_p .$$

It is possible some rows of Ψ_θ to be almost linear dependent. In these cases the explained above regularization can be applied, which is also presented in [1]. Appropriate way to determine the low limit $\sigma_{\min}$ in the Gauss-Newton modification is to compute it periodically. In this case the dimension of $\sum_r$ depends on the current values of $\sigma_{\min}$ and σ_i. The advantage is that introducing a limit on the smallest singular value in $\sum_r$, the updating directions in which the problem is ill-conditioned are skipped. The Gauss-Newton modification starts with $\sigma_{\min}$, large enough to provide numerically stable behavior of the optimization procedure. After minimizing the cost function, the corresponding estimates $\hat{\theta}^h_{\mathrm{opt}}$ should be closer to the optimal ones than the initial estimates. Next step is to decrease $\sigma_{\min}$, which leads to more ill-conditioned matrix $\sum_r$. But approaching the optimum, the procedure has better behavior for ill-conditioned problems. However, with decreasing of $\sigma_{\min}$, the presence of uncertainties in the observations may cause an increase of $J_{\hat{\theta}^h}$. Therefore the optimization procedure stops when $\min J_{\hat{\theta}^h}$ increases and the optimal estimates correspond to $\sigma_{\min}$ that provides the best fit.

3.2.2. Modifications for problematic and missing data

Important requirements imposed on ASIC is to handle with lack of data and incorrect records. In some cases the apriori information is not enough to correct the problematic datasets. So the procedures repairing the datasets are followed by modifications of the estimation algorithms that account for the presence of the remaining uncertainty in the data. To deduce the algorithms we took the idea of WLS, where the residuals are taken into account in J_θ with different weights. This approach is applied to all estimators. The weights depend on the number of available data at the corresponding time instant. For every MISO model a diagonal weighting matrix W is introduced with elements w_{ii} equal to 1, if all records at the i-th moment are available and less than 1, if part of the input-output data is missing or not precisely determined during the preprocessing. If no apriori information is known about these observations, then a suitable way to compute the weights is to decrease w_{ii} proportionally to the number of the absent input data. If y at i-th time instant is not available, w_{ii}= 0. The last rule is introduced because in this case the difference between the system and the model output cannot be formed. Thus the numerically stable solution of the modified LS, which takes into account the presence of wrong or missing data, is

(6) $$\hat{\theta} = V_{L,1}\Sigma_{L,r}^{-1}U_{L,1}^{\mathrm{T}}Ly\ .$$

Here $L = W^{1/2}$ and the matrices $V_{L,1}$, $\Sigma_{L,r}$ and $U_{L,1}$, are determined by SVD, applied to Ly.

Also, the updating rule (5) in the iterative procedure of the PE methods is modified to account for the missing data. It becomes

$$\hat{\theta}^{h+1} = \hat{\theta}^h - \mu^h U_{L,1}\Sigma_{L,r}^{-1}V_{L,1}^{\mathrm{T}}Le_p\ .$$

The matrices $V_{L,1}$, $\Sigma_{L,r}$ and $U_{L,1}$ are determined by applying SVD to $\Psi_\theta L$ and performing the regularization technique explained above.

3.3. Validation criterion and determination of the models types and structures

To measure the quality of the models outputs we use a modification of Variance Accounted For (VAF) that is [11]

(7) $$\mathrm{VAF} = \max(0,\ 1-\mathrm{var}(e)/\mathrm{var}(y))\times 100\%\ .$$

This function takes values between 0% and 100%. It increases when the output predicted by the model approaches the measured output. In ASIC we account the case of wrong or missing data not just during the parameters estimation, but also in the models validation. Here the criterion, denoted by VAF_L , depends on the considered above weights in the following way

$$\mathrm{VAF}_L = \max(0,\ 1 - \mathrm{var}(Le)/\mathrm{var}(Ly)) \times 100\% \,.$$

This criterion is more sensitive to time instants with full input-output records in contrast to the moments with absent or incorrect data. The criterion modification is necessary since when any data is missing, (7) can decrease not because the model is inappropriate, but because some missing factors are not precisely determined. This causes incorrect output forecast.

To determine the appropriate models types and structures, the datasets are divided into two parts. First part containing 2/3 of the available data is used for parameters estimation and the remaining 1/3 part – for validation. For each dataset related to a given product, the set of estimators is run for a set of models with different structures. The models' set is defined by the upper limits of the polynomials' dimensions. They are assigned before the run of ASIC. The criterion VAF_L is computed for each model and the MISO representation with maximum VAF_L is chosen.

4. Market system

To present the abilities of ASIC for multivariate system identification we investigate part of the market, which is the relation between the retailers and the customers. The retailers use the demand models to increase their profit, by optimizing their actions. Usually these models are static representations of the connection between the retailers' actions and the market reaction which is the unit sales. The actions are the prices, discounts, ads and displays of a given product and its cross-related products.

Attempting to find an adequate demand model, the researchers point their attention at the dynamic character of the market behaviour. For instance, a dynamic model for investigation of the effect of the shocks effect is applied in [12]. It is used in the analysis of the quality of the stock market and to investigate the asymmetric shocks behaviour. A dynamic demand model is also used in [5] for different commodity groups (food, alcoholic drink and tobacco, clothing and footwear, energy, etc.).

The sales behavior caused by a promotion is a dynamic process. It is usually connected with an initial overstocking during the promotion and later, as an effect of this overstock people start buying less than usual, despite the promotional price. This effect cannot be accounted by a static model. For that reason dynamic models are used to represent the system behavior. Thus our aim is to find a MIMO dynamic demand model, which will be used to forecast units' sales from a selected group of products. The products are collected into categories to decrease the system inputs under consideration.

The investigated system is subject to different kinds of uncertainties. Usually they have time-varying character, such as the competitors' activities, weather and social effects and so on. Also the products relations that are not accounted in the multivariate structure lead to uncertainty, which is usually time-varying. As many factors affect the market dynamics and no relevant apriori information is available, it is not possible to make preliminary assumptions about the concrete mathematical representations. In the supermarkets, hypermarkets or chains of stores, there are thousands of products. This makes impossible to apply manual actions during the identification. Hence it is completely necessary the models determination to be realized automatically.

Another important characteristic of the considered system is the time period for data collection. Usually the observations are stored weekly, so it is not appropriate to apply special experiments for the purpose of identification, as they would continue for a long time. Moreover, such experiments can lead to large economic losses. This means that the available information has to be used in spite of the possible problems with the datasets.

4.1. Approaches for market dynamics representation

In this paper we compare two kinds of dynamic models. The first kind is a combination of a static MISO and a dynamic SISO parts mentioned in Subsection 4.1.1. The other approach is related to a dynamic MISO demand model described in the next subsection.

A big variety of static demand models exists. The static MISO parts are based on the following models. The multiplicative model [13] has the form

$$y_k = e^{\alpha_0} u_{1,k}^{\alpha_1} u_{2,k}^{\alpha_2} \dots u_{m,k}^{\alpha_m} e^{e_k}, \text{ for } \alpha_i < 1. \tag{8}$$

Here α_i are the model parameters and e_k is the residual. Taking the logarithm of both sides of (8), the log-log model [16] is obtained given with the following linear relation:

$$\ln y_k = \alpha_0 + \alpha_1 \ln u_{1,k} + \alpha_2 \ln u_{2,k} + \dots + \alpha_m \ln u_{m,k} + e_k. \tag{9}$$

This model is widely used in practice to represent the effects of the prices, discounts, advertisements and other factors on the sales behavior [6, 13]. Problems may arise if $u_{i,k}$ or y_k is zero (then $\ln 0 = -\infty$) or less than zero. To avoid this disadvantage it is possible such observations to be replaced by a small positive value. Another alternative is to replace $u_{i,k}$ in (8) with $e^{u_{i,k}}$. The resulting representation is the general exponential model

$$y_k = e^{\alpha_0} (e^{u_{1,k}})^{\alpha_1} (e^{u_{2,k}})^{\alpha_2} \dots (e^{u_{m,k}})^{\alpha_m} e^{e_k}, \text{ for } \alpha_i < 1. \tag{10}$$

Taking the logarithm of both sides of (10) leads to the semi-log model which is again a linear relation:

$$\ln S_k = \alpha_0 + \alpha_1 x_{1,k} + \alpha_2 x_{2,k} + \dots + \alpha_m x_{m,k} + e_k, \text{ for } \alpha_i < 1. \tag{11}$$

Both demand models estimated by ASIC are dynamic. But the way of accounting for the processes in the models is a combination between (9) and (11). The sales and the prices are logarithmized, but the discounts (that can be neglected) and the other retailers' actions are not modified.

The concrete data is related to chains of stores where the products type and the possible types of promotions are known in advance. Thus it is possible to formulate assumptions about the units' prices and sales. To avoid the problems with negative or zero observations, we imposed additional restrictions on the data. They are connected with the following possible cases and assumptions:

- Zero sales are due to the lack of demand, the lack of products in the warehouses or to database errors. As the products quantities in the stores and in the warehouses are not taken into account in the demand models, it is assumed that the zero sales are errors. So these values should be corrected.

• Negative sales appear when the products returned from the customers are more than the sold ones or to database errors. As the products quality (which is the main reason to the returned products) is not taken into account in the model, it is assumed that the negative sales are due to errors and these values should be corrected.

• Zero price is due to a promotion or a database error. As the available datasets are not related to such promotions, these cases are considered as errors and the values should be corrected. If that kind of promotions is applied, then the zero prices should be taken into account (for example by replacing them with small positive values or by modifying the static MISO part).

• Negative price is assumed to be database errors and should be corrected.

The above considerations are implemented in ASIC by the following prices and sales restrictions introduced for all products:

$$\text{price}_k = \begin{cases} \text{price}_{k-1}, & \text{price}_k \le 0 \\ \text{price}_k, & \text{price}_k > 0 \end{cases}, \quad \text{sale}_k = \begin{cases} \text{sale}_{k-1}, & \text{sale}_k \le 0, \\ \text{sale}_k, & \text{sale}_k > 0. \end{cases}$$

Other problematic observations are the outliers, i.e. when the actions remain the same or their deviation is not sizable, but the sales are changing significantly. These situations are due to barcode reading errors or if the data collection is performed for periods that are not exactly one week or if unaccounted promotions are applied only for customers with loyalty cards and etc. Another case is database errors due to technical reasons or errors occurring during the process of data aggregation (when a number of stores is observed). At present the shaving procedure is used to account for such problematic observations, but it repairs only the outliers that are outside the upper and lower limits.

The missing data may be due to database errors, but the usual case is the products to be not on the market for periods of time. Thus to provide appropriate datasets, the empty records again are filled with suitable values. Actually if a competitive to a given product is not on the market, normally the sales of the product increase due to the cannibalization effect. If a compatible cross related product is absent, the sales of the available product decrease, because of the complementary effect. The resulting biases in the sales cannot be determined or appropriately accounted by the demand models. Normally this uncertainty leads to shifted estimates. To obtain in a more adequate way the models parameters, the modification introduced in Subsection 3.2.2 is used. The modified estimators are most sensitive to the cases where the input-output data is full. The sensitiveness decreases when part of the observed processes are not available (the data is not appropriate for estimation of the concrete model with the preliminary chosen multivariate structure).

These are the main problems and the corresponding assumptions and solutions. The two dynamic approaches are considered and compared below.

4.1.1. Demand models containing static MISO and dynamic SISO parts

The automatic approach for obtaining the dynamic part of demand models designed in [7] contains two parts. The first part is a modification of the model of Blatberg-Wisniewski, which is a multivariate static representation. It handles the factors mentioned above gathered by the retailers (Fig. 2) and provides an initial sales forecast. The second part is a SISO dynamic model (more precisely it is a linear regression model), with input – the difference between the forecasted sales, computed by the static model and the preliminary determined baseline product sales. The output of the regression model is a dynamic correction of the unit sales forecast. The aim of the representation on Fig. 2 is to determine if the introduced dynamic part would increase the model accuracy. ASIC designed for identification of the dynamic part of the demand models is based on SISO representations. The automatic identification procedure, which is a generalization of this work, is based on MISO regression models explained below.

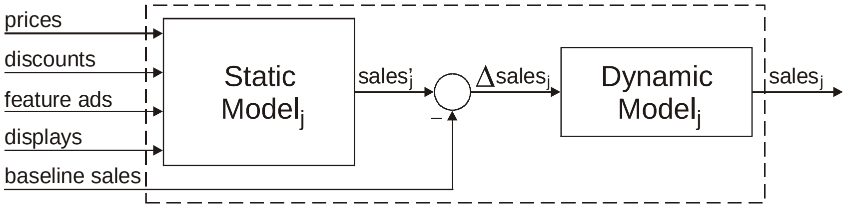

Fig. 2. Model related to the *j*-th product containing static MISO and dynamic SISO parts

4.1.2. Dynamic MISO demand models

The combination of two kinds of models complicates the analysis of the market representation and complicates the application of different techniques from the system theory, such as stochastic observers for optimal sales forecast in the presence of uncertainties in the data and in the market behavior [2]. Another disadvantage of the model is that it represents the whole market dynamics related to a given product by one SISO regression model.

To describe the system behavior in a more natural way and to improve the model accuracy we choose MISO regression models. The inputs of such a model are all factors, related to a concrete product sale and the output is the forecasted sale.

The information about the different features and displays is accounted by two sets of flags. Each flag corresponds to a given advertisement or display activity. Usually there are several kinds of ads and displays related to a product.

Sometimes this leads to enormous size of the parameter vectors. Also the size is very sensitive to the maximum degrees of the inputs polynomials $B_{ji}(q^{-1})$. To decrease $\dim(\theta_j)$, we represent the connections between the flags and the product sales by relations with $nb_{ji} = 1$ (Relations 1 on Fig. 3).

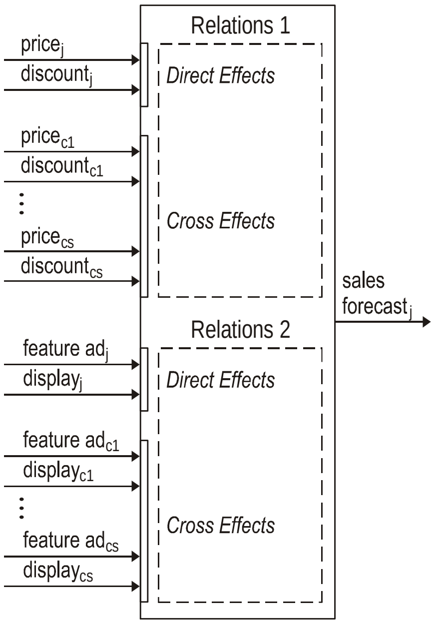

Fig. 3. The *j*-th dynamic MISO product. The indices of the s significant cross related products are denoted as *c*1, *c*2, …, *cs*.

But since the sales dynamics strongly depends on the prices and especially on the discounts, the effect of these factors is modelled by relations with $nb_{ji} \geq 1$ (Relations 2 on Fig. 3).

4.2. Cross effects determination method

As above mentioned, the computational burden and the model accuracy depend on the number of inputs under consideration. Therefore, to improve the ASIC performance and results, we suggest a cross effects determination (CED) procedure, which is an automatic approach, assessing the significant cross effects in multivariate structures. After CED execution, the identification is applied for MISO systems with sub-sets of only significant inputs. We developed the method for the considered market system, but it can be applied for other multivariate systems if the assumptions explained below are fulfilled:

- the main effect of the factors continues on few or several samples,
- the inputs have immediate effect on the output (the delay is one sample for discrete systems),
- the maximums of the input-output cross-correlation functions correspond to one sample delay.

Actually the significant part of the market dynamics is the promotion effect. To avoid the case of overstocking, the retailers run the promotions for two or three weeks, which leads to spikes in the sales for the corresponding periods of time. So the main effect of the input actions is few or several samples. Since the data is collected weekly, practically the second assumption is also fulfilled as the time sample is long enough and the customers' reaction starts during the first week of the promotion. If there is a delay, which is greater than one week a generalization can be made taking into account the place of the global maximum of the correlation functions discussed below.

The idea of CED is based on estimation of the cross-correlation functions between a given output and the set of input processes. After preprocessing the observations, the processes $\tilde{R}_{ij}(\tau)$ are determined, where $i = \overline{1, m}$, $j = \overline{1, l}$ and the lag $\tau \in [1; N-1]$. With $\tilde{R}_{ij}(\tau)$ is denoted the estimate of the cross-correlation function between the *i*-th input $u_{i,k}$ and the *j*-th output $y_{j,k}$. To decrease the influence of the random variations in $\tilde{R}_{ij}(\tau)$, they are filtered by a low-pass filter. Because of the discretization, one sample delay exists between $u_{i,k}$ and $y_{j,k}$, and if they are correlated, then $\tilde{R}_{ij}(1)$ is (or is near) the global maximum of $\tilde{R}_{ij}(\tau)$. Next $\tilde{R}_{ij}(\tau)$ are scaled so that the first value of the processes becomes $\tilde{R}_{ij}(1) = 1$. If $u_{i,k}$ and $y_{j,k}$ are not correlated, the noise component in $\tilde{R}_{ij}(\tau)$ dominates and after scaling, the deviation of $\tilde{R}_{ij}(\tau)$ for $\tau > 1$ increases. After these modifications it is checked if

$$R_{\text{lower}} < \tilde{R}_{ij}(\tau) < R_{\text{upper}} \quad \text{for} \quad \tau \in [a, b] \text{ and } 1 < a < b < N .$$

If the inequality is fulfilled for preliminary chosen *a, b,* R_{lower} and R_{upper} , then it is assumed that the corresponding input has significant influence on the output. Otherwise, the input is considered as not significant and will be neglected. Adjusting the parameters *a, b,* R_{lower} and R_{upper} , we can change the number of significant cross effects.

During the identification, only a subset of inputs is used to estimate the cross effects. As the sales are most sensitive to the discounts, these processes are used for cross effects determination. If the discount of a given product has a sizable effect on the corresponding output, then all retailers actions related to this product are considered as significant and are accounted in the model.

5. Test, results, conclusions and future work

5.1. Test description

In the previous work [7] it was shown that introducing a dynamic part after the static demand model leads to significant increasing of the extended model accuracy. Here two tests are applied. The aim of the first test is to determine whether the modified estimators, explained in Subsection 3.2.2, provide better results than the standard estimators. ASIC generating the MISO demand models is applied. In this test real data are used from a products category, containing 14 products, where all of them are not on the market in different periods of time. The data is collected weekly and the maximum observation interval for modifying the estimators and VAF criterion will be given in more details in an additional paper.

The aim of the second test is the comparison between the two ASIC procedures that generate the demand models (described in 4.1.1 and 4.1.2). The procedures are applied for identification of real market systems, so again we used real datasets. They contain the input-output data related to 1200 products, collected in categories and taken from 65 stores.

The maximum polynomials dimensions are important parameters fixed before the run of both tests. Since there is a requirement the $D(q^{-1})$ polynomials dimensions for MISO-ARARX models to be few times greater than the other polynomials dimensions, in this test it is chosen

$$\max(na_j) = \max(nb_{ij}) = \max(nc_j) = \max(nd_j) = 3,$$

but for MISO-ARARX, $\max(nd_j) = 6$.

To compare the modified and the standard estimators, as well as the two approaches, the corresponding values of VAF_L are computed and compared. Also (for the second test) the applicability of each estimator is assessed. Such aposteriori information can be used to determine the set of estimators that take part in ASIC, if it is run again for similar datasets.

5.2. Results, discussions and conclusions

The approach based on static MISO and dynamic SISO parts is denoted as “I approach” and the dynamic MISO approach as “II approach”. In Table 1 the average values of VAF_L are compared for both standard and modified estimators, obtained by II approach.

As the sensitivity of the modified estimators decreases when the data is not full (the data is not appropriate for the chosen multivariate model structure), the bias in the parameters estimates decreases. The conclusion from the result in the table is that the models obtained by the modified estimators that account for the presence of unrealistic or missing records are more accurate compared to the models obtained by the standard estimators.

Table 1. Criterion values for models related to products from a category

Validation criteria / Demand model	Standard estimators	Modified estimators
	$VAF_{L,II}$	$VAF_{L,II}$
Product 1	66.88	74.57
Product 2	71.00	74.19
Product 3	80.49	81.01
Product 4	77.94	80.01
Product 5	68.08	69.61
Product 6	73.48	76.41
Product 7	76.83	78.15
Product 8	74.80	78.00
Product 9	78.29	81.91
Average and VAF_L	74.20	77.10

The results from the second test are shown below. The average values of for both approaches are:

$$\text{average}(\text{VAF}_{L,I}) = 31\%\,,\ \text{average}(\text{VAF}_{L,II}) = 65,2\%\,,$$

where VAF_L are obtained for I approach and $\text{VAF}_{L,I}$ are corresponding to II approach.

The test results are also presented on Fig. 4 and Table 2. The figure includes VAF_L for all representations, obtained by the two approaches. To observe the improvement of the models adequateness, when MISO regression models are used, the values are sorted. Fig. 4 shows that the MISO dynamic models are significantly better than the models with dynamic SISO part. It is also obvious from the average values of the criterion given above. This result is in accordance with the discussion in Subsection 4.1.2.

Table 2 presents the estimators' applicability for the concrete market systems. The percentage ratios give a posteriori information about the applicability of the estimators that take part in the two approaches. This result can be used to decrease the computation time. For instance if the approaches are executed again for the same or similar products categories, it is suitable to remove IV and EMLS from ASIC as they are most appropriate for less than 5% of the cases. Also EMLS is an iterative procedure, which increases considerably the computation time.

If the values of $\max(n_{aj})$, $\max(b_{ji})$, $\max(nc_j)$and $\max(nd_j)$ grow, VAF_L will increase, but the computational time will also increase significantly. This is especially true if thousands of products are accounted for.

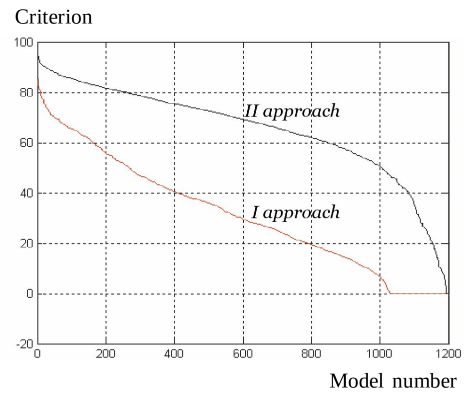

Fig. 4. Sorted values of VAF for all models:
$\mathrm{VAF}_{L,I}$ – I approach, $\mathrm{VAF}_{L,II}$ – II approach

Table 2. The number of most appropriate models, obtained by each estimation procedure in percent

Method	Applicability [%]	
	I approach	II approach
LS	4.2	6.4
GLS	8.7	58
EMLS	3.8	4.3
WLS	12.1	8.9
IV	4.7	3.1
PE-ARMAX	30.1	32.8
PE-OE	25.2	20.2
PE-BJ	11.2	18.5

A disadvantage of the second approach is the great number of the models' parameters, which makes it much slower. As it was mentioned, this is another reason to choose low maximum values of the polynomials' orders. A way to avoid this problem is discussed below.

5.3. Future work

A disadvantage of both identification procedures is the enormous computation time for models generation (especially for the second approach). To deal with this problem recursive estimators for SISO regression representations [8] were developed. They were used to update the dynamic parts of the demand models generated by I approach. A similar set of recursive procedures was designed for the MISO regression representations. It is used for weekly update of the demand models generated by II approach. Additional procedure checks whether the sets of cross effects accounted in the MISO models are still appropriate for the current system behavior. If for a given set, new cross effects appear or other become insignificant, then the structure of the corresponding MISO model is updated. After that ASIC is applied only for the corresponding model, but not to all MISO models. As the market system is time-varying, appropriate modifications of the standard recursive estimators are applied.

The state space representation of the market behavior is appropriate for the considered multivariate system, since usually the number of inputs and outputs is tremendous. A good solution is to apply subspace methods for MIMO demand models. Also the resulting models will be further specified for example by state space optimization based on a project gradient search [17]. As this is an iterative approach, it is possible to arise a problem with the initial conditions. The combination of subspace identification and PE state space methods is an appropriate way for automatic MIMO models generation.

To improve the sales forecast, we will develop Kalman Filters (KF). If aposteriori information about the constraints upon the observed processes is available, then Extended KF (EKF) [2] would be designed.

If the promotion effects are compared with the regular cases, it is easy to conclude that the system behavior varies significantly. Therefore Interacting Multiple Models (IMM) estimators [3] would be suitable to account the time-varying nature of the market.

Acknowledgements. The authors would like to thank the company Retail Analytics Ltd. for providing the real data used in the experimental part of the paper.

R e f e r e n c e s

1. B j o r k b e r g, J., G. K r i s t e n s s o n. Three-Dimensional Subterranean Target Identification by Use of Optimization Techniques. – In: Electromagnetics Research, PIER **15**, 1997, 141-164.
2. B i s h o p, G., G. W e l c h. An Introduction to the Kalman Filter. – In: Siggraph – Course 8, 2001.
3. B l a c k m a n, S. S a m u e l. Multiple Target Tracking with Radar Applications. London, Artech House, 1986.
4. C o s t a, M. A., A. P. B r a g a, B. R. M e n e z e s. Improving Generalization of MLPs with Sliding Mode Control and the Levenberg-Marquardt Algorithm. – Neurocomputing, September 2006.
5. D e s c h a m p s, J. P h i l i p p e. Full Maximum Likelihood Estimation of Dynamic Demand Models. – J. of Econometrics, **82**, 1997, 335-359.
6. D i j k, A. Studies in Local Marketing. PhD thesis. ISBN 90-5335-024-1. The Netherlands, Labyrint Publications, 2004.
7. E f r e m o v, A. An Automatic Identification Cycle for Dynamic Demand Model Generation. International Conference on Computer Systems and Technologies. – In: International Conference on Computer Systems and Technologies – CompSysTech, 2005, Vol. 5, 1-6. **http://anp.tu-sofia.bg/aefremov/index.htm**
8. E f r e m o v, A. Recursive Estimation of Dynamic Time-Varying Demand Models. – In: International Conference on Computer Systems and Technologies – CompSysTech, 2006. **http://anp.tu-sofia.bg/aefremov/index.htm**
9. E l d e n, L. Numerical Linear Algebra and Applications in Data Mining. Preliminary Version. Department of Mathematics, Linkoping University, Sweden, December 20, 2005.
10. G a r i p o v, E. System Identification. Part I. Introduction and Part II. Identification by Discrete Stochastic Regression Models. Second Revised Edition. Technical University of Sofia, 2004.
11. G o m e z, J. C., A. J u t a n, E. B a e y e n s. Wiener Model Identification and Predictive Control of a pH Neutralisation Process. – In: IEE Proc. – Control Theory and Applications, Vol. **151**, Issue 3. ISSN: 1350-2379, May 2004, 329-338. **http://publish.uwo.ca/~ajutan/**

12. G o n z a l o, J e s u s, O s c a r M a r t i n e z. Large Shocks vs. Small Shocks. (Or Does Size Matter? May Be So.) – J. of Econometrics, **135**, 2006, 311-347.
13. H a n s s e n s, D. M., L. J. P a r s o n s. Handbooks in Operations Research and Management Science. Vol. **5**, Chapter 9 – Econometric and Time Series Market Response Models. J. Eliasberg & G. Lilien, Eds. Elsevier Science Publishers, 1993.
14. M a, C., L. J i a n g. Some Research on Levenberg-Marquardt Method for the Nonlinear Equations. – Applied Mathematics and Computation, Vol. **184**, Issue 2, 2007, 1032-1040.
15. M a t s u b a r a, M., Y. U s u i, S. S u g i m o t o. System Identification and Order Determination Using Canonical-Forms. – Intern. J. of Innovative Computing. Information and Control, ISSN 1349-4198, Vol. **1**, September 2005, No 3, 527-545.
16. S a u n d e r s, J. The Specification of Aggregate Market Models. – European J. of Marketing, Vol. **21**, 1987, 5-47.
17. V e r d u l t, V. Non-linear System Identification – A State Space Approach. PhD thesis. Twente University Press, The Netherlands, 2002. **http://doc.utwente.nl/38684/**
18. V e r h a e g e n, M., V. V e r d u l t. Filtering and System Identification. Lecture Notes for the Course sc4040 (et4094) Edition. Delft Center for Systems and Control, Delft University of Technology, 2003.
19. V e r h a e g e n, M., V. V e r d u l t. Filtering and System Identification: An Introduction to using Matlab Software. Software Manual for the Course sc4040 (et4094) edition. Delft Center for Systems and Control, Delft University of Technology, June, 2003.