International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:4, 2010

System Identification Based on Stepwise Regression for Dynamic Market Representation

Alexander Efremov

Abstract—A system for market identification (SMI) is presented. The resulting representations are multivariable dynamic demand models. The market specifics are analyzed. Appropriate models and identification techniques are chosen. Multivariate static and dynamic models are used to represent the market behavior. The steps of the first stage of SMI, named data preprocessing, are mentioned. Next, the second stage, which is the model estimation, is considered in more details. Stepwise linear regression (SWR) is used to determine the significant cross-effects and the orders of the model polynomials. The estimates of the model parameters are obtained by a numerically stable estimator. Real market data is used to analyze SMI performance. The main conclusion is related to the applicability of multivariate dynamic models for representation of market systems.

Keywords—market identification, dynamic models, stepwise regression.

I. Introduction

AN accurate representation of the relation between the retailers actions and the sales reaction predetermines the efficiency of many retailers activities. The estimation of the future demand is using by the retailers to make decisions regarding their promotion strategy, shelf-space allocation, products display, advertisements, coupon systems, etc. The paper is focused on the identification of market systems, more precisely supermarket and hypermarket chains. The specifics of these systems are analyzed in order to choose appropriate mathematical representations as well as the data mining techniques, which have to be included in SMI.

The reaction of investigated market systems is naturally dynamic especially with respect to factors like promotions and advertisements. In general, the customers performance is changing in time after applying of promotions. Normally, the initial reaction (an increase of product sales) is followed by a decreased demand, due to the process of overstocking, or loosing of customers interest. Possibly this is one of the reasons for the increased number of attempts within the last years to obtain more accurate demand representations by using of dynamic models [1], [2], [3]. To present the effect of modeling the dynamic aspect of the market, the accuracy of static and dynamic models, generated by SMI is observed in section V-B. The experiment is based on real data, published in Internet (see subsection V-A).

Usually the hypermarket chains consist of hundred thousands of products. Furthermore a set of actions is associated with each product. Hence the overall input-output dimension of these multiple input multiple output (MIMO) systems is extremely large. It is impossible to apply manual actions during the identification, so the process of obtaining an overall demand model should be performed without human intervention. This is one of the requirements imposed on SMI.

The problem for single MIMO model estimation becomes unfeasible when the input-output dimension is huge. Thus the system should be decomposed into subsystems, which are considered independently during the identification. For that reason SMI generates sets of multiple input single output (MISO) static or dynamic demand models (see section II). Each set of models represents a particular product category.

Another important specific of the market system is the non-linear relation between the dependent and the independent characteristics. Different models are used [4] in order to account for this non-linearity. One of the most frequently used models are the linearly parameterized representations (such as log-log and semi-log models). They are linear with respect to the parameters but the relation between the dependent characteristic and the factors is in general non-linear. The transformations, applied to the data are mentionned in subsection III-C.

A. Efremov is with the Department of Production Automation, Faculty of Automatics, Technical University of Sofia, Bulgaria.
For contacts see http://anp.tu-sofia.bg/aefremov/index.htm.

Manuscript received December 31, 2009; revised April, 2010.

II. Market Models

To make the modeling process feasible, when the input-output data is related to tremendous number of products, the system is decomposed into product categories. Let a given category contains r products and m retailers actions are associated with these products. MIMO static and dynamic representations of the relation between the dependent variable and the available factors are considered below. The MIMO static regression model is

$$y_k = Bq^{-1}u_k + e_k, \tag{1}$$

where $u_k \in \mathcal{R}^m$, $y_k \in \mathcal{R}^r$ and $e_k \in \mathcal{R}^r$ are vectors containing the input, output data and the residual respectively at the k-th time instant. The matrix $B \in \mathcal{R}^{r\times m}$ contains the model parameters. The operator q^{-i} introduces a delay of i time periods (i.e. $q^{-i}u_k = u_{k-i}$). Usually the time period of the considered systems is one week. The introduced delay of one period in (1) is necessary as the market system is naturally discrete. Therefore the effect of the retailers actions on the product sales is observed at least one week after the execution of u_k.

One of the simplest dynamic regression models, applied widely in practice is the Auto-Regressive model with eXogenous input (ARX). The MIMO-ARX model is

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

where $\mathrm{A}(q^{-1}) \in \mathcal{R}^{r\times r}_{n_a}(q^{-1})$ and $\mathrm{B}(q^{-1}) \in \mathcal{R}^{r\times m}_{n_b}(q^{-1})$ are polynomial matrices. To simplify the explanation will be

assumed that all polynomial orders are $n_{a,\ell i} = n_{b,\ell j} = n$ for $\ell, i = \overline{1,r}$ and $j = \overline{1,m}$. The polynomial matrices in (2) can be defined as

$$\mathrm{A}\left(q^{-1}\right) = I_r + A_1 q^{-1} + \ldots + A_n q^{-n},$$
$$\mathrm{B}\left(q^{-1}\right) = 0_{r\times m} + B_1 q^{-1} + \ldots + B_n q^{-n}.$$

The first term in $\mathrm{B}(q^{-1})$ is a zero matrix to account for the one week delay, as discussed earlier. One way to perform the modeling process is to estimate simultaneously the parameters of all polynomials in $\mathrm{A}(q^{-1})$ and $\mathrm{B}(q^{-1})$. The models are linear with respect to the parameters. Therefore the parameters can be easily separated from the regressors and their optimal estimates to be determined analytically. There are two ways to obtain such presentation from the above MIMO models. One way is all parameters to be gathered in a matrix. As y_k is a vector, the regressors should be collected in a vector as well [5]. The main disadvantage of this representation is that the explanation of each component of y_k is obtained by the same set of factors and the polynomials degrees should be $n_{a,\ell i} = n_a$ and $n_{b,\ell j} = n_b$. Hence non-significant independent characteristics and unnecessarily high model orders would be used to represent some components of y_k. The mentioned requirement about the model structure is not practically grounded. The other way to present the MIMO system is to combine the model parameters in a vector, but the regressors to be collected in a matrix [6]. Here the above disadvantage is avoided. The set of factors representing each dependent variable and the polynomials degrees can vary. This makes the second approach more suitable for identification of MIMO linear systems.

In practice, the number of products in a single category may grow to few thousands [7], which is in fact the dimension of y_k. On the other hand, the vector u_k has a few or several times larger dimension than y_k. This is the reason to decompose each product category into MISO systems. To obtain their representations, the models (1) and (2) are divided into r MISO models of the form

$$y_{\ell,k} = \mathrm{B}_{\ell.}\left(q^{-1}\right) u_k + e_{\ell,k} \tag{3}$$

for static and

$$\mathrm{A}_{\ell.}\left(q^{-1}\right) y_k = \mathrm{B}_{\ell.}\left(q^{-1}\right) u_k + e_{\ell,k} \tag{4}$$

for dynamic models. $\mathrm{A}_{\ell.}(q^{-1})$ and $\mathrm{B}_{\ell.}(q^{-1})$ are the ℓ-th rows of $\mathrm{A}(q^{-1})$ and $\mathrm{B}(q^{-1})$. Let p is the total number of parameters/factors in the last two models and N is the number of observations. The models (3) and (4) can be presented in a universal form by collecting the regressors in a vector $\varphi_{\ell,k} \in \mathcal{R}^p$ and the parameters in a vector $\theta_\ell \in \mathcal{R}^p$. At the k-th time instant ($k = \overline{1,N}$) for (3) and (4) holds

$$y_{\ell,k} = \varphi_{\ell,k}^T \theta_\ell + e_{\ell,k}. \tag{5}$$

The intercept is omitted, because of the preprocessing technique, discussed in subsection III-E. To simplify the next expressions the index ℓ of the current MISO model will be skipped. As it is seen from (3), the factors of a MISO static model are the p own and cross products retailers actions $u_{j,k-1}$ (for $j = \overline{1,m}$, $p = m$), applied in the beginning of each time period. For dynamic models of type (4) from order n, the regressors are the previous negative own and cross product dependent variables $-y_{i,k-d}$ (for $i = \overline{1,r}$), and the previous actions $u_{j,k-d}$ (for $j = \overline{1,m}$). The time shift is $d = \overline{1,n}$. Hence the total number of factors in (4) is $p = n(m+r)$. Note that SWR algorithm determines a subset of significant factors and the real number of factors p' is in general less than or equal to m for model (3) and $p' \leq n(m+r)$ for model (4). Regardless of the models type (static or dynamic), the representation (5) for the whole data sample can be written as

$$y = \Phi\theta + e. \tag{6}$$

The vectors $y, e \in \mathcal{R}^N$ and the data matrix $\Phi \in \mathcal{R}^{N\times p}$ are

$$y = [y_1 \ \ldots \ y_N]^T,\ e = [e_1 \ \ldots \ e_N]^T,\ \Phi = [\varphi_1 \ \ldots \ \varphi_N]^T.$$

This notation is used in the next sections, to present the implemented numerically stable estimator and the SWR procedure.

III. Data Preprocessing

The appropriate data preparation before the modeling stage may increase significantly the models accuracy. The preprocessing techniques used in SMI are discussed below.

A. Short Sample Window and Uninformative Factors

The first restriction imposed on the data is connected with the length of the sample window. All products, which are on the market for a period, less than N_{min} (the parameter N_{min} is assigned before to run SMI) are not included in the demand model. An important property of the signals is the persistency of excitation. If for instance a given factor is constant or almost constant for the observation interval, it is not appropriate for modeling. A persistency of excitation check is applied to all available factors. The inappropriate factors are removed.

B. Missing and Unrealistic Records

It is normal the data samples to contain missing records. Furthermore, as the multivariable models account for cross-products relations, all observations associated with a category has to be aligned with respect to the weekly data points. Usually the first/last week of appearance of the products on the market is different. Hence an additional source of missing data is the alignment of the observations.

First step in the discussing preprocessing technique is to determine the baseline prices and sales. These processes are derived from the available prices and sales. To obtain the baseline price $p_{b,k}$ for a given product at the k-th time instant, a weighted moving average is applied to the price p_k. The resulting process $p_{f,k} = \frac{1}{N_{\alpha,k}} \sum_{i=-c}^{c} \alpha_{k+i} p_{k+i}$ (where $N_{\alpha,k} = \sum_{i=-c}^{c} \alpha_{k+i}$) is used in the recursive rule

$$p_{b,k} = \begin{cases} p_k, & \text{for } |p_{f,k} - p_k| < |p_{f,k} - p_{b,k-1}|, \\ \max\left(p_k, p_{b,k-1}\right), & \text{otherwise.} \end{cases}$$

The weights α_k are formed such that the spikes in the price (mostly due to promotions) are penalized with a lower weight. The baseline sales $s_{b,k}$ are determined by using of the

weighted moving average $s_{f,k} = \frac{1}{N_{\beta,k}} \sum_{i=-c}^{c} \beta_{k+i} s_{k+i}$. The weights β_k are formed in a similar way as α_k, but with respect to the spikes in the sales. The process $s_{f,k}$ is used in the rule

$$s_{b,k} = \begin{cases} s_k, & \text{for } |s_{f,k} - s_k| < |s_{f,k} - s_{b,k-1}|, \\ \min\left(s_{f,k}, s_{b,k-1}\right), & \text{otherwise.} \end{cases}$$

Once $p_{b,k}$ and $s_{b,k}$ are obtained, the unrealistic or missing records can be approximated. In these cases it is assumed that the regular actions are applied. More precisely the corresponding price is replaced with $p_{b,k}$, the sales with $s_{b,k}$, the discount is set to zero and it is assumed that no specific ads and displays are applied on the market.

C. Data Transformation

There are different ways to account for the non-linear relation between the dependent and the independent characteristics. Different non-linear transformations were applied on the data and the resulting models accuracy was checked. The transformation, which maximizes the model precision is the natural logarithm applied on the products' prices. The results in subsection V-B are obtained by using of this transformation.

D. Weight Factor Determination

All values, which are introduced on the place of the unrealistic or missing records have to be taken with a lower weight, during the determination of the model parameters. For this reason a weight factor w_k is used. To determine w_k, an additional vector $w_y \in \mathcal{R}^N$ and matrix $w_f \in \mathcal{R}^{N \times p}$ are introduced. The k-th element of w_y is zero if the dependent y_k before the preprocessing stage is unrealistic or missing (keep in mind that the index ℓ is omitted and $y_k \in \mathcal{R}$). Otherwise $w_{y,k} = 1$. Similarly the element $w_{f,ki} = 0$, if the value of the i-th factor at the k-th time instant is introduced by the logic, discussed in subsection III-B. Otherwise $w_{f,ki} = 1$. A possible way to determine the weight factor by using of w_y and w_f is

$$w_k = \left(\frac{w_{y,k}}{p} \sum_{i=1}^{p} w_{f,ki} \right)^2 \tag{7}$$

If the dependent and all factors at the k-th time instant are available in the initial dataset, $w_k = 1$. On the other hand, if y_k and/or all factors are not available in the initial dataset, w_k is zero. And finally, if y_k is not missing or unrealistic, but some of the entries of φ_k are added by the procedure, explained in subsection III-B, w_k is between zero and one. The weight factor decreases with decreasing the number of reliable values in φ_k. A further penalty of the records, when φ_k contains values, introduced by the SMI algorithm, is obtained by raising the expression on the right hand side of (7) on power 2.

E. Data Normalization

The final data preprocessing in SMI is to normalize all factors and the dependent variable as

$$\varphi_{i,k} \leftarrow \frac{\varphi_{i,k} - \bar{\varphi}_i}{\sigma_{\varphi_i}}, \text{ for } i = \overline{1,p} \text{ and } y_k \leftarrow \frac{y_k - \bar{y}}{\sigma_y}, \tag{8}$$

where $\bar{\varphi}_i$ and σ_{φ_i} are the weighted mean and standard deviation of the i-th factor and $\bar{y}$ and σ_y are the dependent weighted mean and standard deviation.
After the performance of SWR, the resulting model parameters have to be denormalized. As a result, the parameter vector of the final model, which can be used to forecast the demand, has to be expanded with an additional term, which is the intercept.

IV. Modeling Stage

A. Parameter Estimation

The weighted least squares (WLS) method is used to estimate the model parameters. The cost function $f(\theta)$, minimized in WLS is the weighted residuals sum of squares. Let $w \in \mathcal{R}^N$ contains all weights w_k and the matrix $W = \mathrm{diag}(w)$. Using the notation introduced in section II, the criterion can be written as

$$\min_{\theta} f(\theta) = \min_{\theta} ||y - \Phi\theta||^2_W.$$

The estimate of the parameter vector, which minimizes $f(\theta)$ can be computed as

$$\hat{\theta} = \left(\Phi^T W \Phi\right)^{-1} \Phi^T W y. \tag{9}$$

The covariance matrix is the inverse of the information matrix $\Phi^T W \Phi$. The matrix inversion may lead to significant loss of precision if Φ has linear or almost linear dependent columns. To decrease the effect of the possible multicollinearity, the matrix $W^{1/2}\Phi$ is decomposed by the economic Singular Value Decomposition (SVD)

$$W^{1/2}\Phi = U \Sigma V^T. \tag{10}$$

The square diagonal matrix Σ contains the singular values σ_i, arranged in descending order

$$\overline{\sigma} = \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p = \underline{\sigma}.$$

To limit the numerical errors due to the inversion, a subset of only significant σ_i is chosen and the decomposition becomes

$$W^{1/2}\Phi = [U_1 \;\; U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T. \tag{11}$$

Thus, using SVD, the estimated vector $\hat{\theta}$ becomes

$$\hat{\theta} = V_1 \Sigma_1^{-1} U_1^T W^{1/2} y. \tag{12}$$

This equation is used in the backward elimination of SWR, considered in the next section.

B. Stepwise Linear Regression

This method [9] employs series of F-tests to check the significance of the factors sequentially added to a linear regression model. In addition the significance of all factors already in the model is re-examined, once a new factor has been added. If non-significant factors are detected, they are removed from the model.
The modeling stage of SMI is connected with repeatedly run of the SWR procedure. Furthermore SWR is iterative and is related to an examination of a set of models. To increase the efficiency of SWR, the covariance matrix is partitioned

into sub-matrices and a technique for inversion by partitioning [8] is applied in the forward selection step. This technique provides a way for calculation of the parameters estimates without a direct matrix inversion at each iteration. In this way the computational burden decreases drastically.

The following three sets of factors are used below to present the SWR algorithm. S_i is the set of i factors, taking part in the current model. S_i^{+j} is the set obtained by adding the j-th factor $\varphi_{j,k}$ to S_i (S_i^{+j} contains $i+1$ factors). S_i^{-j} is the set obtained by removing of $\varphi_{j,k}$ from S_i (S_i^{-j} has $i-1$ factors). The three ANOVA measures: the total sum of squares (SST), the regression sum of squares (SSR) and the residual (error) sum of squares (SSE) are used for the calculation of two F-ratios. They are necessary to assess the significance of the factors at each step of SWR. (Note that the ANOVA measures depend on w, as the weight factor is introduced in the identification scheme.) Let the current model has i factors and the j-th (not entered) factor is tested for significance. The residual and regression mean squares are

$$\mathrm{MSE}_{S_i^{+j}} = \frac{1}{N'-i-1}\,\mathrm{SSE}_{S_i^{+j}}, \quad \mathrm{MSR}_{S_i^{+j}} = \frac{1}{i+1}\,\mathrm{SSR}_{S_i^{+j}},$$

where $N' = \sum_c w_k$. Let also $\mathrm{TIISS}_{S_i^{+j}} = \mathrm{SSR}_{S_i^{+j}} - \mathrm{SSR}_{S_i}$ is the type II sum of squares for the case, when add the j-th factor to a model with i factors. For linear regression models the ratios $F_{\mathrm{o}} = \frac{\mathrm{MSR}_{S_i^{+j}}}{\mathrm{MSE}_{S_i^{+j}}} = \frac{N'-i-1}{i+1}\frac{\mathrm{SSR}_{S_i^{+j}}}{\mathrm{SSE}_{S_i^{+j}}}$ and $F_{\mathrm{p}} = \frac{N'-i-1}{1}\frac{\mathrm{TIISS}_{S_i^{+j}}}{\mathrm{SSE}_{S_i^{+j}}}$ have theoretical F-distribution with $(i+1, N'-i-1)$ and $(1, N'-i-1)$ degrees of freedom (keep in mind that the intercept is omitted). F_{o} and F_{p} are the overall and the partial F-ratio respectively. F_{o} is used in the statistical analysis to test the hypothesis $\theta = 0_{i+1}$ where $0_{i+1} \in \mathcal{R}^{i+1}$ (i.e. if there is a linear relation between the factors and the dependent variable). F_{p} is used to determine whether a given factor has to be added or removed from the current model.

From section II it is easy to see that the static model (1) is a particular case of the dynamic model (2). For this reason SWR algorithm is explained below for the dynamic case. Figure 1 presents the basic idea of this modeling technique.

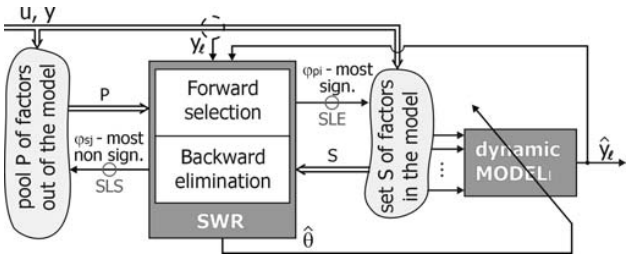

Fig. 1. SWR procedure for the ℓ-th MISO dynamic model.

Step 1. *Initialization*

In order to determine the first (single factor) model, the most significant factor $\varphi_{s_1,k}$ has to be isolated. For all factors the candidate models

$$y = \Phi_{S_0^{+j}}\theta + e_j \text{ for } j = \overline{1,p} \tag{13}$$

are investigated. At this step all unknown models parameters are $\theta \in \mathcal{R}$ and the data matrices are $\Phi_{S_0^{+j}} \leftarrow \Phi_{.j}$. With $\Phi_{.j}$ is denoted the j-th column of Φ. The partial F ratios

$$F_{\mathrm{p}}^{+j} = \frac{N'-1}{1}\frac{\mathrm{TIISS}_{S_0^{+j}}}{\mathrm{SSE}_{S_0^{+j}}}, \text{ for } j = \overline{1,p} \tag{14}$$

are computed. Here the type II sum of squares is $\mathrm{TIISS}_{S_0^{+j}} = \mathrm{SSR}_{S_0^{+j}}$. The p sets S_0^{+j}, related to the candidate models, contain only one factor. Let $\varphi_{s_1,k}$ results in $F_{\mathrm{p}}^{+s_1}$, which is the maximum partial F-ratio and hence the corresponding p-value $p_{\mathrm{p}}^{+s_1}$ is the minimal one. A significance level to enter is used in SWR to restrict the number of factors entering in the model. It is defined with the parameter $\mathrm{SLE} \in [0,1]$ ($\mathrm{SLE} = 0.05$ by default). If $p_{\mathrm{p}}^{+s_1} \le \mathrm{SLE}$, the related model with parameter estimate $\hat{\theta}$ is assumed to be the current best model. Otherwise the SWR procedure is terminated.

Step 2. *Forward selection*

Let the current best model contains i factors with indexes s_1, s_2, ..., s_i. The set of entered factors in the model is S_i. For all $p-i$ remaining (not entered) factors are investigated the models

$$y = \Phi_{S_i^{+j}}\theta + e_j \text{ for } j = \overline{1,p},\ j \neq s_1, s_2, ..., s_i. \tag{15}$$

The unknown parameter vectors are $\theta \in \mathcal{R}^{i+1}$ and the data matrices are $\Phi_{S_i^{+j}} \leftarrow [\Phi_{.s_1}\ ...\ \Phi_{.s_i}\ \Phi_{.j}]$. The set of factors of the j-th model is S_i^{+j}. For each candidate model, the partial F-ratios

$$F_{\mathrm{p}}^{+j} = \frac{N'-(i+1)}{1}\frac{\mathrm{TIISS}_{S_i^{+j}}}{\mathrm{SSE}_{S_i^{+j}}}, \tag{16}$$

are calculated. The type II sum of squares is $\mathrm{TIISS}_{S_i^{+j}} = \mathrm{SSR}_{S_i^{+j}} - \mathrm{SSR}_{S_i}$. Let at the current step of SWR, the factor $\varphi_{s_{i+1}}$ introduced in the model leads to the maximum value $F_{\mathrm{p}}^{+s_{i+1}}$ from all $p-i$ partial F-ratios. If the corresponding p-value is $p_{\mathrm{p}}^{+s_{i+1}} \le \mathrm{SLE}$, the related model is assumed to be the current best model.

Step 3. *Backward elimination*

A normal situation is to exist a certain level of multicollinearity between the factors. Hence after entering a new factor in the model, the significance of the previously entered factors is changing. Therefore after the forward selection, the significance of all factors should be re-investigated. To examine the informativeness of the entered factors, all possible models, obtained by removing of one factor, are estimated. Let the initial model in the backward elimination contains $i+1$ factors. Then the candidate models with i factors are

$$y = \Phi_{S_{i+1}^{-s_j}}\theta + e_j \text{ for } j = \overline{1,i+1}, \tag{17}$$

where the parameter vectors in this case are $\theta \in \mathcal{R}^i$ and the data matrices are $\Phi_{S_{i+1}^{-s_j}} \leftarrow \left[\Phi_{.s_1}\ ...\ \Phi_{.s_{j-1}}\ \Phi_{.s_{j+1}}\ ...\ \Phi_{s_{i+1}}\right]$. The set of variables of the j-th model is $S_{i+1}^{-s_j}$. For each candidate model, the partial F-ratio

$$F_{\mathrm{p}}^{-s_j} = \frac{N'-i}{1}\frac{\mathrm{TIISS}_{S_{i+1}^{-s_j}}}{\mathrm{SSE}_{S_{i+1}^{-s_j}}} \tag{18}$$

is calculated. Here $\mathrm{TIISS}_{S_{i+1}^{-s_j}} = \mathrm{SSR}_{S_{i+1}} - \mathrm{SSR}_{S_{i+1}^{-s_j}}$. Let at the current step of SWR the factor φ_{s_j} leads to the minimum value $F_{\mathrm{p}}^{-s_j}$ of the partial F-ratios. A significance level to stay is used in SWR to restrict the number of factors leaving the model at the backward elimination step. This restriction is defined with the parameter $\mathrm{SLS} \in [0, 1]$ ($\mathrm{SLS} = 0.05$ by default). If the corresponding p-value is $p_{\mathrm{p}}^{-s_j} \geq \mathrm{SLS}$, the related model with a reduced dimension is assumed to be the current best model.

The steps 2 and 3 are repeating until the set of factors out of the model contains only non-significant factors and all entered factors in the model are significant.

There is an additional stopping rule, which is used in SWR. It is introduced to avoid a generation of overfitted models. Different criteria for selection of the model structure are available, which account for the dimension of $\hat{\theta}$. Popular criteria in statistics are the Akaike Information Criterion (AIC) [10], [11], the Bayesian Information Criterion (BIC) or Schwarz Criterion, the Minimum Description Length (MDL) [10], [12], Final Prediction Error (FPE) [13], etc. These criteria include an additional penalty of the cost function, and have extremum with respect to $i = \dim(\hat{\theta})$. A further increase of $\dim(\hat{\theta})$, after the extremum, is connected with possible model overfitting.

The mentioned criteria provide reliable assessment of the overfitting for large datasets ($N \gg 1$). But usually the sample windows of the investigated market systems have short length (keeping in mind that the observations are collected weekly). There are modifications of the above criteria for short datasets, such as corrected AIC ($\mathrm{AIC^c}$) and corrected MDL [10]. With increasing of N, $\mathrm{AIC^c}$ converges to AIC, which makes $\mathrm{AIC^c}$ appropriate for moderate and large datasets as well. The criterion $\mathrm{AIC^c}$ is used in SMI as an additional restriction on the number of entered factors in the models. It has the form

$$\mathrm{AIC^c} = \frac{1}{N'} \mathrm{SSE}\, e^{2\frac{i+1}{N-i-2}} \tag{19}$$

The iterative SWR procedure is terminating when $\mathrm{AIC}^c_{S_i^{+j}} \geq \mathrm{AIC}^c_{S_i}$, where the above criteria are obtained from two successive and consecutive forward selection steps. The model with i factors is chosen as a final model. If a factor is dropped from the model during a backward elimination step, the above rule is not checked and SWR continues.

V. SMI Performance

A. Data Description and Experimental Scenario

The performance of SMI is investigated by experiments with a real-life data. For this purpose a database of a large western supermarket chain, the Dominick's Finer Foods, is used. The data is provided by the James M. Kilts Center, GSB, University of Chicago. It is published at the University's website[1]. The overall dataset contains information about 29 product categories throughout the stores in the chain. The products in the dataset are distinguished by their unique universal product code (UPC). The data consists of weekly observations of the sales, retail prices and an indication of the promotion activities. A subset of four product categories collected from sixteen stores are used in this paper. The chosen sample covers almost four-year period (199 weeks).

SMI is used for generation of two sets of models, based on this data sample. The first set contains MISO static regression models of type (3). The second set is of MISO dynamic models of type (4). Both experiments are run under the same conditions. The thresholds SLE and SLS, discussed in the previous section are $\mathrm{SLE} = \mathrm{SLS} = 0.15$. These levels are not so restrictive with respect to the informativeness of the entered and dropped factors from the models. The main restriction in the SWR algorithm, imposed on the number of factors is related to the possible overfitting of the models, discussed in the previous section.

The minimum models order, providing the ability to represent both aperiodic and oscillating processes is 2. For this reason the maximum dynamic models order is $n = 2$.

The following measures of the models accuracy are observed. The coefficient of determination R^2 is a measure of the closeness between the models and the real system performance. It is computed as

$$\mathrm{R}^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}. \tag{20}$$

In the next subsection R^2 and the adjusted coefficient of determination

$$\mathrm{R}^2_{adj} = 1 - \left(1 - \mathrm{R}^2\right) \frac{N' - 1}{N' - i} \tag{21}$$

are shown. As it was mentioned earlier the ANOVA measures depend on the weight factor w. In this case R^2 and R^2_{adj} are more accurate estimates of the models accuracy, as they are most sensitive with respect to the data when all input-output observations are available. On the other hand their sensitivity decreases, if some factors are missing or have unrealistic values. In the limited case, when all factors and/or the dependent are not initially available, R^2 and R^2_{adj} are not sensitive to the subsequently introduced values by SMI. These statistics and the overall F-ratio are presented in the next subsection.

B. Results

The average values of the statistics for all stores, included in the data set are shown on table I. A graphical comparison of the models accuracy, obtained by SMI is given on figure 2.

TABLE I
AVERAGE STATISTICS FOR THE STATIC AND DYNAMIC MODELS

Product category	Average values for all stores in the data set		
	$\bar{p}$	$\bar{\mathrm{R}}^2$	$\bar{\mathrm{R}}^2_{adj}$
static models			
Cereals	4.98	0.21	0.20
Cheeses	6.08	0.20	0.19
Frozen Entrees	3.70	0.12	0.11
Refrigerated Juices	6.39	0.17	0.15
dynamic models			
Cereals	10.00	0.41	0.38
Cheeses	11.74	0.35	0.31
Frozen Entrees	7.16	0.22	0.20
Refrigerated Juices	12.13	0.32	0.28

[1]The Dominick's database is publicly available at the University of Chicago website: http://research.chicagogsb.edu/marketing/databases/dominicks/

To obtain more readable plot, the models are rearranged with respect to R^2_{adj}, obtained for the static models.

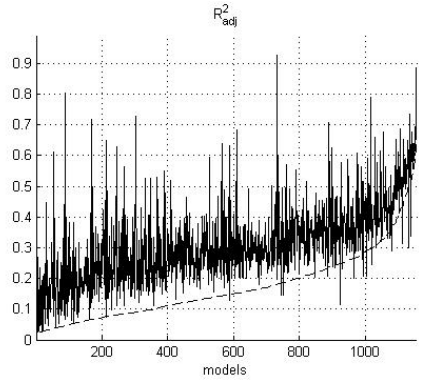

Fig. 2. R^2_{adj} for all static (dashed) and dynamic (solid) models.

Another result, representing the SMI performance is the time, spent for the demand models development. To generate the set of 1155 static models SMI has spent 15 min on a computer system with CPU 2.8GHz, RAM 1GB (SMI is developed on Matlab 7.7). On the other hand the dynamic demand models are obtained for 51 min.

VI. Conclusion and Further Improvements

From the results can be concluded that the SWR algorithm succeeds to gather appropriate sets of factors for the dynamic models. The usage of such models leads to significantly better approximation of the system behavior than the static models. When introduce the dynamic terms in the demand representations, according to R^2_{adj}, the average models precision for all 1155 models increases with 14%.

As it is expected the development time for dynamic models is greater (3.4 times) than the time spent for the static models. As now-a-day the computers have more than one cores, the time for identification can be decreased significantly. This can be obtained by parallelizing the procedures, discussed in the paper. The SWR is the most time consuming procedure. Fortunately the nature of this modeling algorithm allows it easily to be parallelized.

The seasonality is a typical characteristic of the market systems. A further improvement of SMI can be achieved by assessing the possible seasonal component in the data. The same is the situation with the linear trend.

The considered identification is based on the MISO-ARX models. If the residual is a color noise, more appropriate MISO models are ARMAX, ARARX, etc. Better approximation can be achieved by using of prediction error estimators.

In many cases the categories dimension allows to use MIMO regression models. This would further improve the efficiency of SMI. State space identification would be useful as well, when represents the market dynamics, as the system is multivariable.

Acknowledgment

The author would like to thank Hristo Hadjitchonev and Nikolay Marekov from Experian Decision Analytics for their support.

References

[1] P. Deschamps, *Exact Small-Sample Inference in Stationary, Fully Regular, Dynamic Demand Models*. Journal of Econometrics, Volume 97, pp. 51-91, 2000.

[2] J. Gonzalo and O. Martinez, *Large shocks vs. small shocks. (or does size matter? May be so.)*. Journal of Econometrics, Volume 135, pp. 311-347, 2006.

[3] J. W. Hamister and N. C. Suresh, *The impact of pricing policy on sales variability in a supermarket retail context*. Int. J. Production Economics, 2007.

[4] D. Hanssens and L. Parsons, *Handbooks in Operations Research and Management Science, Volume 5*, J. Eliasberg and G. Lilien, Eds. Elsevier Science Publishers, 1993.

[5] S. D. Fassois, *MIMO LMS-ARMAX identification of vibrating structures part I: the method*. Mechanical Systems and Signal Processing, Volume 15, No 4, pp. 723-735, 2001.

[6] M. Leskensa, L. B. M. Van Kessela and P. M. J. Van den Hof, *MIMO closed-loop identification of an MSW incinerator*. Control Engineering Practice, Volume 10, pp. 315-326, 2002.

[7] H. Chen, D. Levy, S. Ray and M. Bergen, *Asymmetric Price Adjustment in the Small*, Journal of Monetary Economics, Volume 55, pp. 728-737, 2008.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing* Third Edition, Cambridge University Press, 2007.

[9] J. O. Rawlings, S. G. Pantula, D. A. Dickey, *Applied Regression Analysis: A Research Tool*, Second Edition. Springer-Verlag New York, Inc., 1998.

[10] F. Ridder, R. Pintelon, J. Schoukens and D. P. Gillikin, *Modified AIC and MDL model selection criteria for short data records*. IEEE Trans. on Instrumentation and Measurement, Volume 54, No 1, pp. 144-150, 2005.

[11] L. Ljung, *System Identification Toolbox. For Use with Matlab*. User's Guide. The MathWorks, Inc, MA, USA, 2004.

[12] P. Grünwald, *The Minimum Description Length Principle*. MIT Press, June 2007.

[13] E. Garipov, *System Identification. Part II - Identification by Discrete Stochastic Regression Models*. Technical University of Sofia, 2004.

[14] M. Verhaegen and V. Verdult, *Filtering and System Identification. A Least Squares Approach*. Cambridge University Press, 2007.

Alexander Efremov received his engineering degree in Systems and Control Engineering in September 1999, and the doctoral degree in production automation with application in the retail industry in February 2008 from the Technical University of Sofia, Bulgaria.

From January 2000 he started his career in the industry in Gallus Ltd as a control engineer, where he made a research on real-time identification for time-varying systems and adaptive control with application in the chemical industry. From August 2000 to May 2003 he was a control engineer in SkyGate BG Ltd (at present RaySat Inc) where he worked in the field of satellite communication. His activities were related to modeling of multivariable nonlinear systems, Kalman filtering, interacting multiple model approach, etc. for satellite direction tracking. In the period December 2004 – May 2008 he was a researcher and scientific developer in Retail-Analytics Ltd. His research was focused on the market systems identification and demand forecast for the retail sector of industry. Currently he works as a product analyst at Experian Decision Analytics. His activities are related to investigation of different ways for modeling of the human performance. The focus is on stepwise linear/logistic regression, dynamic, time-varying representations, decision trees, discriminant analysis and other techniques for scoring and tracking of the human behavior.

Meanwhile, he started his academic carer in 2004 as a part time assistant in the University of Veliko Tarnovo, Bulgaria. From 2005 to 2008 he was an assistant professor of the Department of Programming and Computer System Application of the University of Chemical Technology and Metallurgy – Sofia, Bulgaria. Since 2006 he is an assistant professor at the department of Industrial Automation of the Technical University of Sofia.