# Recursive Estimation of Dynamic Time-Varying Demand Models

Alexander Efremov

***Abstract:*** *The paper presents an implementation of a set of recursive algorithms and their modifications for estimation of the dynamic part of a price model. The reasons for real time model estimation of the sales dynamics are considered. As the market behaviour varies a number of modifications are applied to keep the estimations' sensitivity with respect to the current market behaviour. Also problems with applying the considered procedures for model updating are discussed and solutions are presented.*

***Key words:*** *Recursive Estimation, Demand Model, Time-Varying Systems, Sales Forecasts.*

## INTRODUCTION

The considered demand model is a mathematical relation between the prices, discounts, ads and displays of a given product and of its cross related products. The updated models are designed to forecast product sales from a selected group of products. Usually the products are collected into product categories. In a given store or a group of stores such division of the whole set of products is necessary for decreasing the model's parameters dimension in reasonable limits.

The models discussed in the paper compute the forecasted sales into two steps. First step is to estimate the future sales based on static models. The obtained estimates are used as an initial forecast of the unit sales. The static part handles the main factors gathered by the retailers (such as product price, ads, displays, discounts...). For the initial sales forecast is used a modification of the model of Blattberg-Wisniewski. The second step of the forecast provides an improved representation of the market behaviour, taking into account the dynamic aspect of the market. This step is a correction of the estimated sales using dynamic models. For the second step regression models are used with input – the difference between the forecasted sales, computed by the corresponding static models and the baseline product sales. The output of the regression models is the corrected unit sales forecast.

This paper is focused on the real time estimation of the second part of the demand models. The recursive model updating is necessary because ordinarily are observed thousands of products and it makes the identification procedure too slow. Also after the model estimation could be applied other actions, such as optimal sales forecast and price optimization, so the time for model updating has to be decreased. The recursive procedures can not be adjusted manually for a great number of products. This is a problem because the models quality strongly depends on the choice of procedures parameters maintaining the estimators' sensitivity for time-varying systems dynamics.

As the considered system is naturally time-varying (the market behaviour depends on a non-stationary environment and the sets of products and cross effects are not fixed too), the problem with estimators' sensitivity is investigated and appropriate solutions are presented.

## RECURSIVE ESTIMATORS AND THEIR MODIFICATIONS FOR DYNAMIC DEMAND MODELS UPDATING

The presented algorithms use information about the model structure and parameters that are obtained by an automatic system identification cycle [3]. This procedure is based on running of a set of block-methods for a set of models with different orders. A validation criterion is used for the best model determination. The recursive approach presented below is a continuation of this block-identification procedure.

**Recursive estimators**

For updating of the dynamic part of demand models the following algorithms are applied [1, 2]:

- Recursive least squares (RLS)
- Recursive general least squares (RGLS)
- Recursive extended least squares (RELS)
- Modification of the recursive extended least squares method based on a posterior error (MRELS)
- Recursive extended matrix least squares (REMLS)
- Recursive forecasting error (RFE) for ARMAX models

Part of the procedures estimates the same type of regression models. These estimators are compared and a set of the most appropriate estimators is determined. In the last subsection ARMAX models are estimated by RFE method.

*Recursive Least Squares*

RLS estimates the parameters of ARX model

$$A_k(q^{-1})y_k = B_k(q^{-1})q^{-d}u_k + e_k, \qquad (1)$$

where $u_k$, $y_k$ and $e_k$ are the input, output and the residual in moment $k$; $d$ is time delay and polynomials $A_k(q^{-1})$ and $B_k(q^{-1})$ are of orders $na$ and $nb$ respectively. The vector form of (1) is

$$y_k = \varphi_k^T\theta_k + e_k, \qquad (2)$$

where

$$\theta_k = \begin{bmatrix} a_{1,k} & a_{2,k} & \dots & a_{na,k} & b_{1,k} & b_{2,k} & \dots & b_{nb,k} \end{bmatrix}^T,$$

$$\varphi_k = \begin{bmatrix} -y_{k-1} & -y_{k-2} & \dots & -y_{k-na} & u_{k-d} & u_{k-d-1} & \dots & u_{k-d-nb+1} \end{bmatrix}^T.$$

are the vector of model parameters and the regression vector respectively.

The algorithm consists of the following steps:

1. Computing the error based on $\hat{\theta}_{k-1}$: $\varepsilon_k = y_k - \varphi_k^T\hat{\theta}_{k-1}$
2. Kalman gain determination: $G_k = P_{k-1}\varphi_k(1+\varphi_k^T P_{k-1}\varphi_k)^{-1}$
3. Updating the parameter vector: $\hat{\theta}_k = \hat{\theta}_{k-1} + G_k\varepsilon_k$
4. Updating the covariance matrix: $P_k = (I - G_k\varphi_k^T)P_{k-1}$

The initial conditions of the recursive procedure are:

- Initial parameters estimation is chosen to be $\hat{\theta}_0 = \theta_0^*$, where $\theta_0^*$ is an estimation obtained by the identification cycle for data sets with maximum length $N_0$
- Initial regression vector is constructed based on the last available data
- Initial value of the covariance matrix is $P_0 = \left[\sum_{i=0}^{N_0}\varphi_i\varphi_i^T\right]^{-1}$

*Recursive General Least Squares*

If the residual is a colour noise (in this case it will be denoted by $e_{c,k}$), the above algorithm provides biased estimations. To avoid this problem, different modification of RLS can be applied.

In RGLS, additional forming filter is applied to take into account the colour dynamics of $e_{c,k}$. The system behaviour is presented by ARARX model

$$A_k(q^{-1})y_k = B_k(q^{-1})q^{-d}u_k + \frac{1}{D_k(q^{-1})}e_k. \qquad (3)$$

Here $e_{c,k} = \dfrac{1}{D_k(q^{-1})} e_k$. The vector form of the regression equation is

$$y_k = \varphi_k^T \theta_k + e_{c,k} ,$$

with

$$\theta_k = \begin{bmatrix} a_{1,k} & a_{2,k} & \dots & a_{na,k} & b_{1,k} & b_{2,k} & \dots & b_{nb,k} \end{bmatrix}^T ,$$

$$\varphi_k = \begin{bmatrix} -y_{k-1} & -y_{k-2} & \dots & -y_{k-na} & u_{k-d} & u_{k-d-1} & \dots & u_{k-d-nb+1} \end{bmatrix}^T .$$

The above equation can be written as

$$y_{f,k} = \varphi_{f,k}^T \theta_k + e_k , \tag{4}$$

with

$$\varphi_{f,k} = \begin{bmatrix} -y_{f,k-1} & -y_{f,k-2} & \dots & -y_{f,k-na} & u_{f,k} & u_{f,k-1} & \dots & u_{f,k-nb+1} \end{bmatrix}^T ,$$

$$y_{f,k} = D_k(q^{-1}) y_k , \quad u_{f,k} = D_k(q^{-1}) u_k .$$

The vector form of filter's regression equation is

$$e_{c,k} = \varphi_{d,k}^T \theta_{d,k} + e_k . \tag{5}$$

The filter's parameter vector and the corresponded regression vector are

$\theta_{d,k} = \begin{bmatrix} d_{1,k} & d_{2,k} & \dots & d_{nd,k} \end{bmatrix}^T$ and $\varphi_{d,k} = \begin{bmatrix} -e_{c,k-1} & -e_{c,k-2} & \dots & -e_{c,k-nd} \end{bmatrix}^T$.

RGLS is reduced to two RLS estimators applied to models (4) and (5). The initial conditions are chosen in the same way as in RLS.

*Recursive Extended Least Squares*

RELS is a method that estimates ARMAX models

$$A_k(q^{-1}) y_k = B_k(q^{-1}) q^{-d} u_k + C_k(q^{-1}) e_k . \tag{6}$$

Here the vector of model parameters and the regression vector are extended with the filter parameters and the values of $e_k$ respectively, i.e.

$$\theta_k = \begin{bmatrix} a_{1,k} & a_{2,k} & \dots & a_{na,k} & b_{1,k} & b_{2,k} & \dots & b_{nb,k} & c_{1,k} & c_{2,k} & \dots & c_{nc,k} \end{bmatrix}^T ,$$

$$\varphi_k = \begin{bmatrix} -y_{k-1} & -y_{k-2} & \dots & -y_{k-na} & u_{k-d} & u_{k-d-1} & \dots & u_{k-d-nb+1} & e_{k-1} & e_{k-2} & \dots & e_{k-nc} \end{bmatrix}^T .$$

The obtained model form is the same as (2), but $\theta_k$ and $\varphi_k$ have different structures. The process $e_k$ represents all uncertainties such as measurement noise and unmodelled system dynamics. If $\hat{\theta}_k$ is not optimal, the system dynamics, which is not accounted in the current model leads to colour residual. To generate the regression vector an estimation of $e_k$ is necessary. In RELS $e_k$ is replaced by $\varepsilon_k = y_k - \varphi_k^T \hat{\theta}_{k-1}$.

*Modification of RELS using a posterior error*

Other variant to construct $\varphi_k$ is to use the posterior error $\varepsilon_{p,k} = y_k - \varphi_k^T \hat{\theta}_k$.

This modification has better performance as $\varepsilon_{p,k}$ is a more precise estimation of $e_k$ than $\varepsilon_k$ used in the standard RELS.

*Recursive Extended Matrix Least Squares*

REMLS is a method that estimates ARARMAX models

$$A_k(q^{-1}) y_k = B_k(q^{-1}) q^{-d} u_k + \frac{C_k(q^{-1})}{D_k(q^{-1})} e_k . \tag{7}$$

To obtain the vector form of the regression equation, the vector of optimal model parameters and the regression vector are extended with the filter parameters and the values of $e_{c,k}$ and $e_k$ respectively, i.e.

$$\theta_k = [a_{1,k} \quad a_{2,k} \quad \dots \quad a_{na,k} \quad b_{1,k} \quad b_{2,k} \quad \dots \quad b_{nb,k}$$
$$c_{1,k} \quad c_{2,k} \quad \dots \quad c_{nc,k} \quad d_{1,k} \quad d_{2,k} \quad \dots \quad d_{nd,k}]^T = [\theta_{ab,k}^T \quad \theta_{cd,k}^T]^T,$$
$$\varphi_k = [-y_{k-1} \quad -y_{k-2} \quad \dots \quad -y_{k-na} \quad u_{k-d} \quad u_{k-d-1} \quad \dots \quad u_{k-d-nb-1} \quad \dots$$
$$\dots \quad e_{k-1} \quad e_{k-2} \quad \dots \quad e_{k-nc} \quad -e_{c,k-1} \quad -e_{c,k-2} \quad \dots \quad -e_{c,k-nd}]^T = [\varphi_{ab,k} \quad \varphi_{cd,k}]^T.$$

The obtained model form is the same as (2), but $\theta_k$ and $\varphi_k$ have different structures again. Here $e_{c,k}$ and $e_k$ are replaced by their estimates $\varepsilon_k$ and $\varepsilon_{c,k}$ computed as

$$\varepsilon_{c,k} = y_k - \varphi_{ab,k}^T \hat{\theta}_{ab,k} \text{ and } \varepsilon_k = \varepsilon_{c,k} - \varphi_{cd,k}^T \hat{\theta}_{cd,k}.$$

*Recursive Forecasting Error Method*

The applied RFE is deducted for ARMAX model estimation. The difference between RFE and RELS is in the regression vector, which (for ARMAX models) is filtered with a forming filter, i.e.

$$\varphi_{f,k} = \varphi_k / C_k(q^{-1}).$$

For constructing the regression vector $\varphi_k$ again the posterior error $\varepsilon_{p,k}$ is used as an estimation of $e_k$.

To avoid obtaining of unstable models during the recursive estimation, an additional procedure [2] for each algorithm is applied. At each time instant a model stability check is applied. If the current model is unstable, the roots of the characteristic polynomial are shrinked in the unit circle.

**Modifications of the Recursive Estimators**

The estimators' sensitivity depends on the covariance matrix $P_k$. If the elements of $P_k$ are too small, the algorithm becomes inertial with respect to the new data and the rate of parameters updating decreases. The following modifications are related to changes in the updating rule of the covariance matrix in such a way that the elements of $P_k$ are big enough to maintain the estimators' sensitivity. For $P_k$ can be written

$$P_k = P_{k-1} - \frac{P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{1 + \varphi_k^T P_{k-1}\varphi_k}. \tag{8}$$

The following modifications are applied for each estimator [2]:

- Modification with adding of a positive definite matrix to the covariance matrix
- Modification with a constant forgetting factor
- Modification with a variable forgetting factor
- Modification with a constant trace of the covariance matrix
- Modification with a constant covariance matrix

*Modification with adding of a constant positive definite matrix to the covariance matrix*

The updating equation of the covariance matrix in this modification is

$$P_k = P_{k-1} - \frac{P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{1 + \varphi_k^T P_{k-1}\varphi_k} + S,$$

if $tr(P_{k-1}) \le trP_{\min}$, where $trP_{\min}$ is the lowest limit of the $tr(P)$. If $tr(P_{k-1}) > trP_{\min}$, the covariance matrix is updated by (8).

*Modification with a constant forgetting factor*

The equation (8) in this modification has the form

$$P_k = \frac{1}{\rho_c}\left[P_{k-1} - \frac{P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{\rho_c + \varphi_k^T P_{k-1}\varphi_k}\right].$$

The factor $\rho_c$ is a positive number, but less then 1. If $\rho_c = 1$, the above equation becomes the same as (8). With decreasing of $\rho_c$, the estimations' sensitivity with respect to data increases, but in this case $\hat{\theta}_k$ is more sensitive to the noise. Also in some cases the procedure can become unstable, what is a disadvantage of this modification.

*Modification with a variable forgetting factor*

The disadvantage of the above modification can be avoided with the variable forgetting factor. The equation (8) in this modification has the next form

$$P_k = \frac{1}{\rho_{v,k}}\left[P_{k-1} - \frac{P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{\rho_{v,k} + \varphi_k^T P_{k-1}\varphi_k}\right].$$

The variable forgetting factor $\rho_{v,k}$ can be determined in different ways. An appropriate updating rule is

$$\rho_{v,k} = 1 - \frac{1-\varphi_k^T G_k}{\sigma_e^2 N_E} e_k^2 . \qquad (9)$$

$N_E$ is the effective observation interval and $\sigma_e^2$ is the variance of $e_k$. One disadvantage of the procedure is that an increase of $e_k$ can be caused by other factors that have an influence on the residual, but not by changes in the system dynamics. In this case the estimation error will increase in spite of the time-invariant character of the system behaviour. The concrete realization uses a floating window for recursive determination of the variance of $e_k$.

*Modification with a constant trace of the covariance matrix*

This modification provides a constant estimators' sensitivity. $P_k$ is updating from

$$P_k = \frac{1}{\rho_k}\left[P_{k-1} - \frac{P_{k-1}\varphi_k\varphi_k^T P_{k-1}}{1 + \varphi_k^T P_{k-1}\varphi_k}\right],$$

where $\rho_k$ is

$$\rho_k = 1 - \frac{\varphi_k^T P_{k-1}^2 \varphi_k}{(1+\varphi_k^T P_{k-1}\varphi_k)\; tr(P_{k-1})}.$$

*Modification with a constant covariance matrix*

In this case $P_k = P_{k-1} = \ldots = P_0$. This modification is simple and useful, but it is sensitive to the choice of $P_0$. If the elements of $P_0$ are too small the procedure will become inertial.

## TEST DESCRIPTION AND RESULTS

The presented above algorithms are tested with a real dataset containing data for a number of 21 products. The information about the prices, discounts, ads, displays and seals is collected weekly. The observation interval is 2 years. The first 84 weeks are used for an initial identification, realised by the mentioned above identification cycle. The maximum models' order was 4, but the maximum order of $D(q^{-1})$ polynomial in ARARX model was 8. The remaining 20 weeks are used for the real-time models estimation. The modification with a variable forgetting factor was used to maintain the estimators' sensitivity. For both initial and weekly updated models are determined the values of the validation criterion Variance Accounted For (VAF) [4] by using of the last 20 weeks. VAFs for both models of each product are compared in Table 1.

A number of experiments (not discussed here) were undertaken to determine the most appropriate estimators and modification. As a result for estimation of ARMAX models RFE is chosen and the modification with variable forgetting factor is applied.

**Table 1.**

| VAF /Initial Models/ | VAF/Updated Models/ |
|---|---|
| 100.00 | 99.97 |
| 100.00 | 99.43 |
| 98.75 | 99.05 |
| 75.52 | 75.71 |
| 69.78 | 70.57 |
| 64.00 | 62.12 |
| 63.96 | 62.01 |
| 63.17 | 65.26 |
| 61.80 | 65.98 |
| 61.14 | 58.92 |
| 60.49 | 58.62 |
| 58.92 | 58.49 |
| 57.95 | 57.67 |
| 54.50 | 53.04 |
| 52.04 | 51.75 |
| 50.91 | 50.87 |
| 49.86 | 49.21 |
| 42.40 | 48.55 |
| 40.67 | 40.82 |
| 40.58 | 48.66 |
| 36.78 | 42.66 |

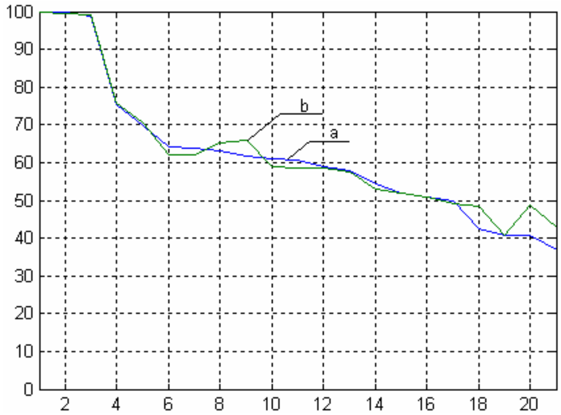


Fig. 1. Sorted values of VAF for the initial models (a) and for the updated models (b)

## CONCLUSIONS AND FUTURE WORK

In [3] were made important conclusions explaining some specific cases. There were considered the reasons for low model accuracy obtained for certain products.

The main advantages of the recursive approach are the ability to assess the time varying system dynamics and also the drastic decrease of the computation time.

There are cases, where the recursive procedures do not provide model improvement. A reason for that could be inappropriate choice of the parameters taking part in (9). It is also possible that optimal model type and structure can be different for that period.

The development of an automatic identification cycle for demand models where the static part is skipped is an object of a future work. In these representations the model input will be extended with the main significant factors gathered by the retailers. A disadvantage of the current demand models is that one regression model is used for representation of the whole market dynamics. Significant improvement is expected if apply multiple inputs single output (MISO) models as the dynamics between each factor and the unit sales will be presented by different dynamic model. The MISO dynamic demand model is a more general and closer to the nature of the considered system.

## REFERENCES

[1] Garipov, E. System Identification. TU–Sofia, 1997.

[2] Velev, K. Adaptive Systems. Sofia, 1995.

[3] Efremov, A. An Automatic Identification Cycle for Dynamic Demand Model Generation. International Conference on Computer Systems and Technologies – CompSysTech' 2005, Varna, Bulgaria.

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

## ABOUT THE AUTHOR

Alexander Efremov, Scientific Developer at Retail Analytics – Bulgaria Ltd.
Phone: +359 02 8163-485, e-mail: alefremov@yahoo.com.