DISCOUNTED WEIGHTED REGRESSION AND FORECASTING

M. Indralingam
Department of Mathematics

DISCOUNTED WEIGHTED REGRESSION AND FORECASTING

M. Indralingam
Department of Mathematics

The recursive methods can be easily adapted to varying parameter regression models. There are two ways of doing this: first using discounting methods and the second is to model the parameter variation explicitly. In this paper discounting method applied to Regression model is described. Effectiveness of this method is demonstrated with simulated data. Forecasting performance with a Real data set is compared with other commonly used regression techniques. A software developed for this research is described. This software can be used in system forecasting.

INTRODUCTION

The recursive estimation procedures introduce an extra dimension to estimation. In addition to the "en bloc" estimation based on the complete data set of N sample points, the analyst is also able to obtain estimates of the parameters for up to N subsets of the data, in a computationally elegant and efficient manner. This is not only useful in estimation of varying parameter models but also can be used in constant parameter models.

The recursive approach to estimation can be traced back to Gauss (1821-1826) although it is also linked with the name Plackett (1950) who rediscovered the results of Gauss and translated them into more useful vector matrix terms. However it was Kalman (1960) who initiated the research with the publication of his paper on " A new approach to linear filtering and prediction problem. The recursive steps computationally consist the following steps:

1. Formulate the model
2. Provide an initial guess or estimate of the model parameters.
3. Forecast the next observation using the most recent estimate of the model parameters.
4. Given the next observation, calculate the forecast error by subtracting the forecast from observation.
5. Update the model parameters by adding a correction term, which is proportional to the forecast error, to each of the parameters.
6. If more data exist return to step 2 or otherwise stop

In this paper , Discounted Weighted Regression (DWR) based on the recursive estimation is introduced. The idea is based on exponential weighted regression (Gilchrist !967) but different discounting factor is used for different parameter. Some what user friendly software is developed in this research. Practitioners involved in system forecasting such as load forecasting, hydrological forecasting can use this software. Moreover we have seen that forecast accuracy of this method is better than the traditional regression models if it is implemented with adequate modification .

RECURSIVE ORDINARY LEAST SQUARE (ROLS) ESTIMATION

Consider a simple regression model

$$Y_t = \underline{X}_t \underline{\theta}_t + e_t \; ; \; t = 1......T \qquad 1.1$$

where $\underline{X}_t = (x_1, x_2, x_3 \ldots x_k)$

$\underline{\theta}_t = (\theta_1, \theta_2, \theta_3, \ldots \theta_t)$

e_t = observation error term with $E(e_t) = 0,\; E(e_t e_s) = 0, E(e_{\cdot t}^2) = \sigma^2$

Then the least square estimation is obtained from minimizing sum of squares:

$$\sum e_{\cdot i}^2 = \sum (Y_i - \underline{X}_i \underline{\theta})^2 \qquad 1.2$$

and the estimator becomes:

$$\underline{\hat{\theta}} = (\sum \underline{X}_{\cdot i}^{/} \underline{X}_{\cdot t})^{-1} . \sum \underline{X}_i^{/} Y_t \qquad 1.3$$

Now suppose this estimator has been computed for t observations then

$\underline{\hat{\theta}}. = \underline{P}_t^{-1} \underline{b}_{\cdot t}$ where $\underline{P}_t = \sum (\underline{X}_i^{/} \underline{X}_i)$, $\underline{b}_t = \sum \underline{X}_i^{/} Y_i$

writing

$$\underline{P}_t^{-1} = \underline{P}_{t-1}^{-1} + \underline{X}_t^{/} \underline{X}_t$$

$$\underline{b}_t = \underline{b}_{t-1} + \underline{X}_t^{/} Y_t \qquad 1.4$$

From this the following recursive relationships can be deduced (Plackett 1950, Young 1984)

$$\begin{aligned} \underline{\hat{\theta}}_t &= \underline{\hat{\theta}}_{-1} + \underline{K}_t v_t \\ v_t &= Y_t - \underline{X}_t \underline{\hat{\theta}}_{t-1} \\ \underline{P}_t &= \underline{P}_{t-1} - \underline{K}_t \underline{X}_t \underline{P}_{t-1} \\ K_t &= \underline{P}_{t-1} \underline{X}_t^{/} (1 + \underline{X}_t \underline{P}_{t-1} \underline{X}_{\cdot t}^{/})^{-1} \end{aligned} \qquad 1.5$$

Also writing

$$w_t = v_t (1 + \underline{X}_t \underline{P}_{t-1} \underline{X}_t^{/})^{-1/2} \qquad 1.6$$

then $$s_t = s_{t-1} + w_{\cdot t}^2 \qquad 1.7$$

where $$s_t = (Y_t - \underline{X}_t \underline{\hat{\theta}}_t)^{1} (Y_t - \underline{X}_t \underline{\hat{\theta}}_t)$$

If an estimator of $\underline{\theta}_k$ is calculated from the first k observations, it can be updated using the set of equations 1.5. There is a clear computational advantages in this formulation in contrast to the "en bloc" method (1.3). It involves no matrix inversion, and all matrix and vector operations that are employed depend on the fixed dimension k (=number of parameters). It also enables the changes in parameters to be tract over time.

PARAMETER VARIATION AND RECURSIVE ESTIMATION

The recursive method described by above algorithm (1.5) can be easily adapted to varying parameter regression models. There are two ways of doing this, first is to apply discounting method and the second is to model the parameter variation explicitly. In this paper we applied the discounting method.

ROLS method use all the information possible by using all the data set. In order to allow for possible parameter variation, it is necessary to remove the effect of 'obsolete' data. The two most obvious procedures for doing this are: first to base an estimation on only the most recent portion of the data and second to weight the data exponentially. We describe here the second procedure which has most theoretical appeal.

Minimizing exponentially weighted sum of squares :

$$\sum e_i^2 = \sum (Y_i - \underline{X}_t \underline{\theta})^2 \alpha^{t-1} \quad \text{where } 0 < \alpha < 1 \qquad 1.8$$

Considering the t observation we can write equations similar to

$$\begin{aligned} \underline{P}_t^{-1} &= \alpha \underline{P}_{t-1}^{-1} + \underline{X}_t' \underline{X}_t \\ \underline{b}_t &= \alpha \underline{b}_{t-1} + \underline{X}_t' Y_t \end{aligned} \qquad 1.9$$

$$\begin{aligned} \underline{\hat{\theta}}_t &= \underline{\hat{\theta}}_{-1} + \underline{k}_t v_t \\ v_t &= Y_t - \underline{X}_t \underline{\hat{\theta}}_{t-1} \\ \underline{P}_t &= \underline{P}_{t-1} - \underline{K}_t \underline{X}_t \underline{P}_{t-1} / \alpha \\ \underline{K}_t &= \underline{P}_{t-1} \underline{X}_t' (\alpha + \underline{X}_t \underline{P}_{t-1} \underline{X}_t')^{-1} \end{aligned} \qquad 1.10$$

The whole idea of including a discounting factor is that estimates respond more quickly to a change in the structure of the model. However this is achieved at the expense of stability. Thus there is trade-off between sensitivity and stability and the usual compromise to choose a value of α which is not far from unity.

Above method can be easily modified to give different discounting factor for different parameters in the model. This takes into account the possibility of different rates of change of parameters in the model. We call this method Discounted Weighted Regression (DWR) Method. Most of the recursive relation can be obtained from

$$\underline{P}_t^{-1} = \underline{B}\,\underline{P}_{t-1}^{-1}\underline{B} + \underline{X}_t'\underline{X}_t \qquad 1.11$$

where $\underline{B} = diag(b_1^{1/2}, \ldots b_k^{1/2})$

then the recursive scheme becomes;

$$\underline{P}_t^{-1} = \underline{R}_t - \underline{K}_t\underline{K}_t'(1 + \underline{X}_t\underline{R}_{t-1}\underline{X}_t') \qquad 1.12$$

$$R_t = \underline{B}\underline{P}_{t-1}^{-1}\underline{B}$$

$$\underline{K}_t = \underline{R}_t\underline{X}_t'(1 + \underline{X}_t\underline{R}_t\underline{X}_t')^{-1}$$

Intuitively discounting the information in this manner has an appeal if we think the matrix $\underline{P}_t$ as proportional to information or precision matrix. In the implementation one has to supply the initial values of parameter with their respective $\underline{P}_o$ values (usually diagonal), discount matrix $\underline{B}$. Note that If $\underline{B} = \underline{I}$ the DWR become ROLS.

DWR FORECASTING SYSTEM.

The system was developed in PASCAL using algorithm described above. It has many facilities such as:

(i) Common Transformation of Variables

(ii) Various plot routines for variables, parameter estimates and recursive residuals

It has been tested with standard regression procedure. For instance if B = I it should give the same parameter estimates as ordinary estimates. In appendix I some program outputs are given.

Simulation study

To see the effectiveness of DWR the following two simple models are considered.

Model 1

$$Y_t = a_t + b_t X_{1t} + c_t X_{2t} + e_t \quad ; \quad e_t \sim N(0, \sigma_*^2) \qquad 1.13$$

Where $X_{1t} \sim N(3,1)$, $X_{2t} \sim N(5.0, 1.0)$, $\sigma^2 = 0.005$

The following different parameter variation is considered

(i) Step variation (Sudden jump)

(ii) ramp variation (increasing)

Plots actual variation and estimated variation is presented in the figure 1. It can be seen that estimated value indicate the variation in the parameter variation. To get the better tracking of the parameter estimate we have to use smoothing procedures (Kalman 1960). H owever, t he f inal e stimate o f the parameters are some what close to actual vales

Model 2

$$Y_t = a_t X_1 + b_t X_2 + e_t \qquad 1.14$$

where $a_t = \beta_t.\exp(-\beta_2.t)$, $b_t = \beta_3(1-\exp(-\beta_2.t))$, $e_t \sim N(0,0.055)$

Estimated parameter and actual parameter variation are given in the figure 2 It shows the reasonable tracking of the variation of parameters.

Empirical Study of Real Data

The Lydia Pinham Vegetable compound advertising has been extensively used by researchers to investigate the different aspects of the sales –advertising relationship. Yearly data from 1907-1960 is provided by Palda (1965). In 1925 the FDA disputed Pinkham's label claiming change in advertising method. In all four periods of advertising method can be delineated:

1907 - 1914 : Universal remedy
1915 - 1925 : Relief for menstrual problem
1926 - 1940 : Vegetable tonic
1941 - 1960 : Same as 1915 – 1926

Palda's Analysis incorporated these shifts in copy the use of dummy variable. Thus incorporating the dummy variable in the following model at various duration time stated above.

$$S_t = a_o + a_t S_{t-1} + a_2 A_t + e_t \qquad 1.15$$

where S_t : Sales at time t.

A_t : Advertising expenditure at time t

Above model is the popular Brand loyal model used in advertising research (Broadbent 1979) in which a_t is known as brand loyal parameter. The DWR estimate of the model is given in the figure 3.

To g et a b etter p reference o f t he s tructural m odel of this nature, instrumental Variable (IV) is used (Maddala 1977). This because of error occurs in the variable in the right hand side. This procedure is widely applied in econometrics In this study we have implemented instrumental variable in recursive manner in DWR We call this procedure DWR-IV. For detail see the appendix 2.

Results

Effects o f dif ferent a dvertising m ethods o n parameter can be seen from the figure.

1907 – 1914 :(sample 1 to sample 8) : Increase of rate of change of brand loyal parameter.

1941 – 1960 (35 to 54): Decreasing rate of change of brand loyal parameter.

Forecast Performance

We use Mean absolute percentage error (MAPE) measure (1.16) to compare the accuracy of the DWR, DWRIV with Ordinary Regression model.

$$MAPE = .\frac{1}{k}\left(\sum\left(\frac{|Y_{t+1} - \hat{Y}_{t+1}|}{Y_{t+1}}\right)\right) \times 100 \qquad 1.16$$

MAPE

Method	1951- 1961	1957 - 1961
Normal Regression	5.5	6.5
DWR	5.74	6.70
DWR-IV	3.7	4.9

Table 1

From the above table1 forecast accuracy of DWR with IV gives better forecast performance than the other. This shows Recursive Estimation procedures with specific modification can be used to enhance the predicting ability.

CONCLUSION

Recursive estimation of regression models which can take into account parameter variation was described in the paper. A prototype version of a DWR software was developed for this research. Practitioners involved in the system forecasting can use this effectively. Most attractive feature is the graphs of parameter estimates over time, which can be used to identify parameter shift . The case study discussed in this paper demonstrates the identification of effects of different advertising methods. Further improvement in the DWR method such as incorporation of smoothing procedures is being investigated.

REFERENCES

PLACKETT , R. L (1950) Some theorems on least Squares, Biometrica, 37, 149-157

KALMAN , R.E (1960). A New Approach to linear filtering and Prediction Problems", Journal of Basic Engineering, Vol. 29, No. 11, 1325-1333

GILCHRIST. W. G. (1967). Methods of Estimation Involving Discounting, J.R. Stat. Soc, B, 29, 355 – 369

YOUNG , P. (1984). Recursive Estimation and Time Series Analysis, An Introduction", Springer-Verkly, Berlin Heidelberg.

PALDA, K. (1965). "The Measurements of Cumulative Advertising Effects". Journal of Business, 38, 162 – 185.

BROADBENT ,S. (1979),"One way TV advertisements works",J. Mkt. Res. Soc, Vol.21,3 139-165

MADDALA, G.S (1977), "Econometrics", Mc Graw-Hill Inc

Figure 1 : Parameter estimates using DWR

Discount B = diag (0.95, 0.95, 0.95)

Int θ = (0.0,0.35,0.05)', P= diag (0.2,0.005,0.007)

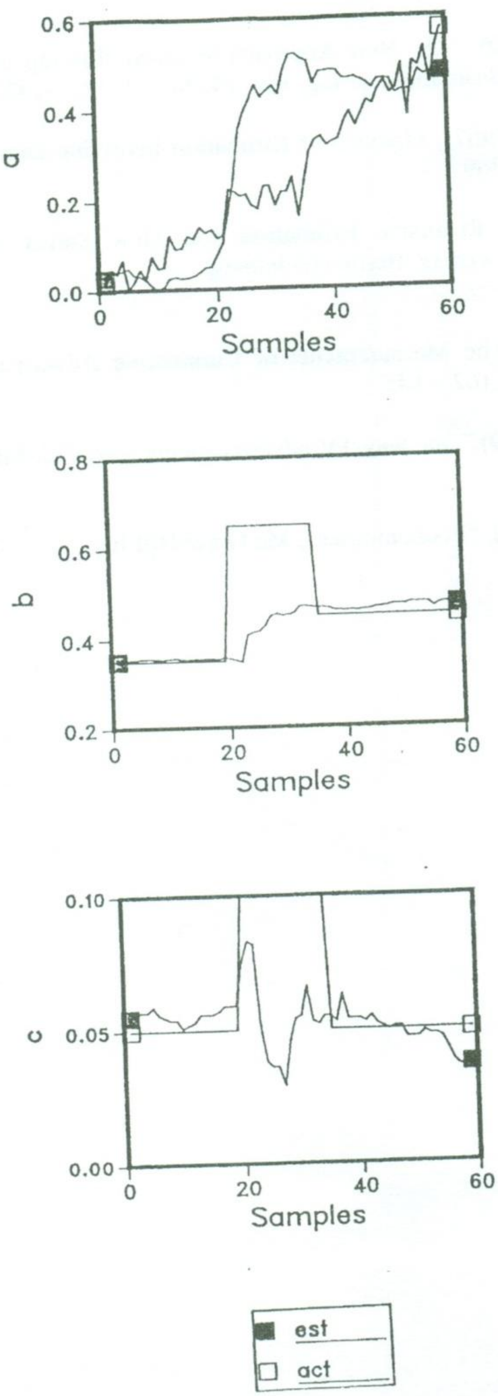
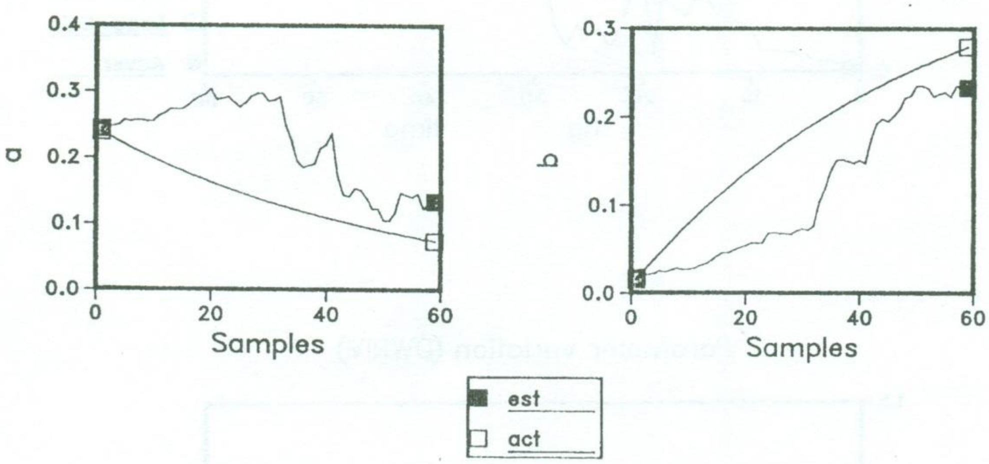


Figure 2 6Parameter estimates using DWR



Parameter variation (DWR)

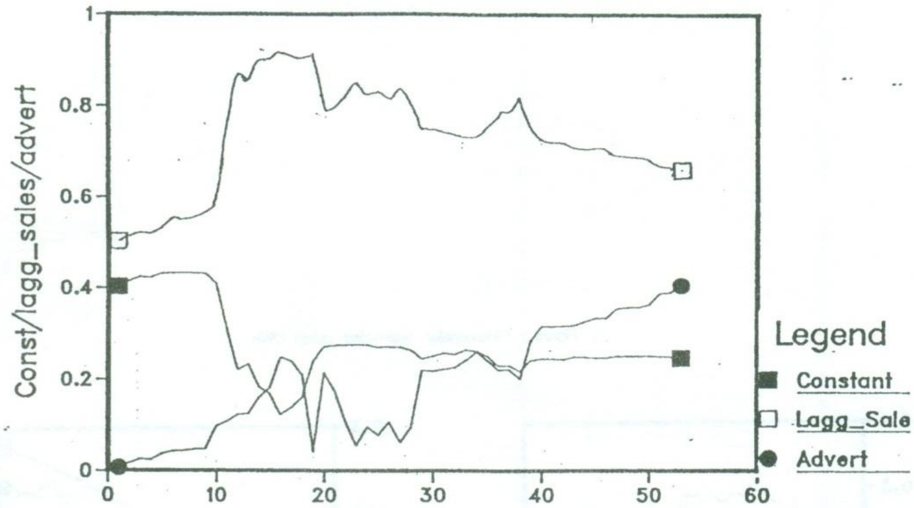


fig 3

Parameter variation (DWRIV)

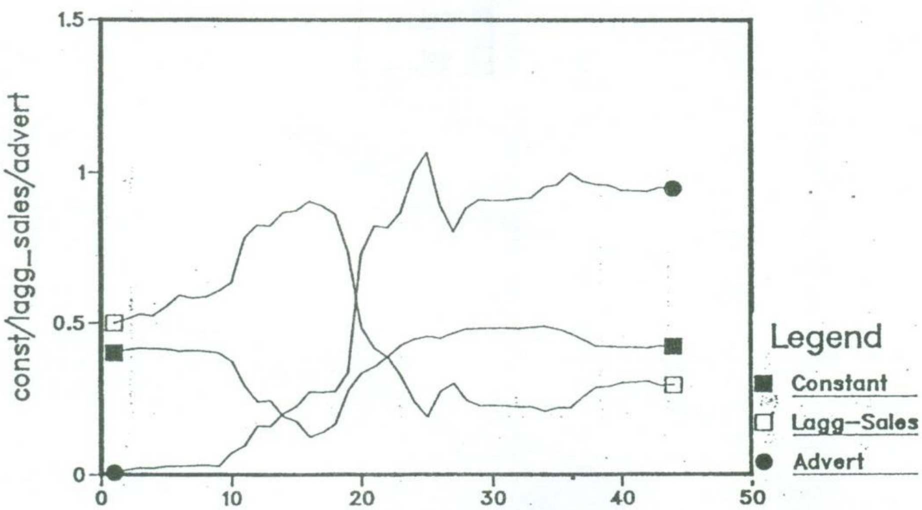


fig 4

APPENDIX I

SCREEN FOR FILE HANDLING

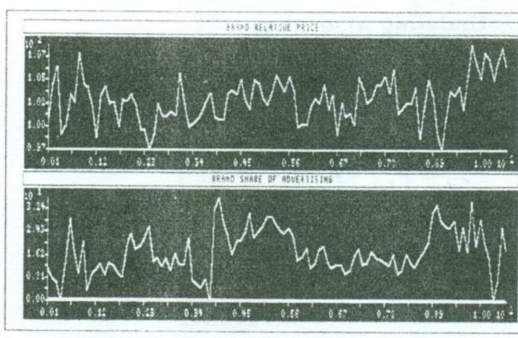
DISCOUNT WEIGHTED REGRESSION PROGRAM

+--------------- File Names ---------------+

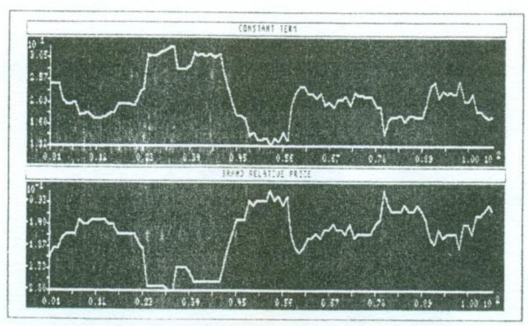
Choose Defaults	Current Settings
A Datafile	c:Test3.dat
B U-Matrix	c:Umatrix3.dat
C Param. estimators	c:Paramete3.dat
D Beta(s)	c:Bmatrix3.dat
E Output file	c:Plot.dat

X Filenames OK

PLOT SCRRENOF VARIABLES



PLOT SCREEN OF PARAMETER ESTMATES



APPENDIX 2

IV procedure Adapted for DWR

For Simple Regression Model:

$Y_t = Z_t.\theta_t + e_t$; where $Z_t = (1, Y_{t-1}, z_t)$ IV steps will be

Step 1 : Initialise with the parameter values obtained from ROLS or other starting values say θ_{t-1}

Step 2 : Use this parameter to find the estimated value of Y_{t-1} using the equation
$\hat{Y}_{t-1} = Z_{t-1} * \theta_{t-1}$

Step 3 : Form instrumental variable vector
$\hat{X}_t = (1, \hat{Y}_{t-1}, z_t)$

Step 4 : Use the following algorithm given to update the parameters

Step 5 : Return to step 2 until the whole data is processed.

Updating Algorithm

$$\underline{\hat{\theta}}_t = \underline{\hat{\theta}}_{-1} + \underline{k}_t v_t$$
$$v_t = Y_t - \underline{Z}_t \underline{\hat{\theta}}_{t-1}$$
$$\underline{P}_t = \underline{P}_{t-1} - \underline{K}_t \underline{Z}_t \underline{P}_{t-1}$$
$$\underline{k}_t = \underline{P}_{t-1} \underline{X}_t^1 (1 + \underline{Z}_t \underline{P}_{t-1} \underline{X}_t^1)^{-1}$$

(More detail can be obtain from the author)