Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

A new insight into the algebraic structure of the exponential smoothing algorithm of Brown

A. Bellacicco
Department of Systems and Organizations Theory, Teramo University, Italy.

Abstract

The paper deals with the problem of the algebraic characterization of the exponential smoothing considering its iterative structure in the parameters estimation process. Exponential smoothing algorithms are well known as forecasting techniques and support the simulation process too. Actually, the forecasting process avoids any reference to an underlying stochastic process and related assumptions. On the other side it is possible to identify two main algebraic components which can give a new status to the exponential smoothing algorithm so that a new meaning can envisaged which can be a suitable tool not only for the forecasting process but also for simulating some potential evolution of a sequence of events. The paper considers the algebraic components of the exponential smoothing algorithm and gives a new strategy for the optimal estimation of the parameters of the model.

1 Introduction

The aim of the paper is to presents the algebraic structure of the exponential smoothing algorithm when a finite number of iterations is already realized. Following the usual way of introducing exponential smoothing algorithms we can write down the classical equation of the exponential smoothing estimator of Brown, [1]and [2], as follows:

$$Y_{t+1} = \alpha Y_t + \beta Y_{t-1} \qquad (1)$$

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

where $\beta=1-\alpha$ and $0\leq\alpha\leq1$.
It is easy to see that the estimation of the parameter α can be realized following some different criteria which are substantially reducible to the choice of an optimization criterion related to the forecasting error [2].
As it is well known, the iteration can be considered as a tool for enlarging the forecasting horizon .Problem we can face is to evaluate the algebraic structure of the set of parameters involved by the iteration process.

2 The iteration process

Iteration process can be considered as a leading feature of the exponential smoothing algorithm. We can now present the first steps of the iteration process in order to evidence the hidden algebraic structure of the parameters of the model.
We can start with (1) and consider a second iteration like :

$$Y_{t+2} = \alpha\, Y_{t+1} + \beta\, Y_t \tag{2}$$

We can insert Y_{t+1} as follows :

$$Y_{t+2} = \alpha\, [\alpha\, Y_t + \beta\, Y_{t-1}\,] + \beta Y_t \tag{3}$$

In the same way expand Y_{t-}as follows :

$$Y_{t+2} = \alpha^2\, Y_t + \alpha\beta\, Y_{t-1} + \beta\alpha\, Y_{t-1} + \beta^2 Y_{t-2}$$

$$Y_{t+2} = \alpha^2\, Y_t + 2\,\alpha\beta\, Y_{t-1} + \beta^2\, Y_{t-2} \tag{4}$$

We can go on considering the case Y_{t+3} which shows a symmetry between exponents and shifts of the time lag as we can see easily :

$$Y_{t+3} = \alpha\, Y_{t+2} + \beta\, Y_{t+1}$$

We obtain easily

$$Y_{t+3} = \alpha^3\, Y_t + 2\alpha^2\beta\, Y_{t-1} + \alpha\beta^2 Y_{t-2} + \beta\, Y_{t+1}$$

Finally we get

$$Y_{t+4} = \alpha^4\, Y_t + 2\,\alpha^3\,\beta\, Y_{t-1} + \alpha^2\,\beta^2 Y_{t-2} + \alpha\beta\, Y_{t+1} + \beta\alpha^2\, Y_t + 2\alpha\beta^2 Y_{t-2} + \beta^3 Y_{t-2} \tag{5}$$

Some comments can be proposed considering the two iterations:

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

i. we can extend the previous results to whatever horizon we like either forward or backward;
ii. there is a symmetry between the power of the two parameters (α,β) and the lags of the variable Y;
iii. we can ascertain an algebraic structure both of the parameters and of the lags;
iv. we never introduced random shocks so that we can ask what happens to (4) and its generalizations when we introduce a random shock

3 Algebraic characterization

This paragraph is devoted to the development of the algebraic rules identified in (4) and into its generalization.
Let us consider equations (4) and (5).We can see that apart from the numerical coefficients there is some similarity between the power sof α and β and the usual binomial expansion.
We can see that equation (4) exactly represents the coupling of two components :

i. the *polynomial* component in α and β;
ii. the *shift* operator component .

The polynomial component,$P_2(\alpha,\beta)$ is characterized by the expansion of $(\alpha+\beta)^2$ while the shift component , $S_2(t)$,shows is effects on the time lag of the variable Y in terms of the sequence of lags starting from the horizon t. We can now state two main asserts :

1.the *polynomial* components can superimposed as far as they satisfy the rules of a multiplicative abelian group with an homomorphism representation on the set of the exponents,[3];

2.the *shift* components can be superimposed as far as they satisfy the axioms of a multiplicative abelian group with an homomorphic representation on the set of the temporal indices, see [3].

As a consequence we can find an isomorphic representation on the set of the exponential functions of the type $e^{n+m} = e^m e^n$.
Some words of explanation are required in order to justify both the used terminology and the two asserts.
First at all we use the term *superposition* in order to emphasize in our case the usual notion of additive effects of a set of actions like for instance the iterative application of the two operators, respectively $P_k(\alpha,\beta)$ and $S_k(t)$.

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

In order to show that the two mentioned operators satisfy the axioms of an additive group it is easy to see that the usual axioms of a multiplicative group are satisfied.
We see respectively:

i. $P_h(\alpha,\beta)\,[P_k(\alpha,\beta)\,] = P_{h+k}(\alpha,\beta)$
ii. $P_h(\alpha,\beta)\;P_0(\alpha,\beta) = P_h(\alpha,\beta)$
iii. $P_h(\alpha,\beta)\,P_{-h}(\alpha,\beta) = P_0(\alpha,\beta)$

In the same way we can write

iv $S_h(t)\;S_k(t) = S_{h+k}(t)$
v. $S_h(t)\;S_0(t) = S_h(t)$
vi. $S_h(t)\;S_{-h}(t) = S_0(t)$

The **0** element is ensured by the absence of any shift in one case and in any use of weights for past data So that the 0 means : no forecasting beyond the the narrow horizon t for Y. The inverse element means a shift on the back side of the instant t.
Problem now is to consider what happens of the * operation :

$$P_h(\alpha,\beta) * S_h(t)\,[Y]\,. \tag{6}$$

We can omit the * symbol when it is clear the contest of the discourse.
We observe that :
i. It is not possible to interchange the occurrence of the two operators;
ii. It is necessary to have the same index h at the bottom;
iii. It is *not* true the distributive property while it true

$$P_{h+k}(\alpha,\beta)[S_h(t)\,S_k(t)] = P_{h+k}(\alpha,\beta)\;S_{h+k}(t) \tag{7}$$

As a general remark we can see that the composite structure can be envisaged as a particular algebraic structure of non commutative operators.
We can now ask what is the usefulness of the previous considerations on the process of forecasting and simulation of a time series.
We can observe that a forecasting process is described by a sequence which gives the time horizon and the length of the series we need for attaining the forecasting horizon.
We can now ask what is the class of all isomorfic operators can be identified where the previous ones can be considered as models so that a more powerful theory on this operator structure can be found.

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

4 More about the stochastic component

We consider a forecasting operator without assuming any random shock on the available time series.
We can ask what kind of random process can be considered in order to say something about the action of the two mentioned operators. We observe that the mentioned operators are able to produce a good forecasting provided an adaptive approach is followed so that the spread of the time series can be catched.,[4]. We propose an hybrid way of estimating the parameter α which is the unique parameter of the forecasting operators.
We consider the following random walk, see[4] :

$$Y_t - \alpha Y_{t-1} = b + \varepsilon_t \tag{8}$$

We see that we can estimate the parameter α as far as the parameter b by the usual regression machinery that is by the mean square error technique.

We can insert the previous equation into the operators of shift so that we have that the time series is transformed as follows.
We can evaluate (α,b) so that equation (1) can be rewritten:

$$Y_{t+1} = \alpha Y_t + \beta Y_{t-1}$$

$$Y_{t+1} = \alpha (Y_t - Y_{t-1}) + Y_t \tag{9}$$

$$\alpha = \frac{Y_{t+1} - Y_t}{Y_t - Y_{t-1}} = \frac{Y_t - Y_{t-1}}{Y_{t-1} - Y_{t-2}} \tag{10}$$

We have now to face a new problem related to the evaluation of the new parameter in terms of the previous way of writing down the shift operator. Shift operator has a status from the algebraic point of of view quite similar to the exponential operator.
We can consider from (5) and (6) the following evaluation. Let us consider equation (8) as a way for estimating α.We can consider (8) still as a random walk and the evaluation of α can be obtained in terms of the average of the sequence of ratios like (8) so that the ratio can be considered as a new operator which will be identified by the symbol $R_{1,1}$ in order to recall that we are relating two lags whose length is 1.,[4] and [5].
We can go back to the equations (1) and (5) and we can say that (1) is just a random walk model as far as α and β are related by the constraint

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

$\alpha+\beta=1$,see [5]. The parameter α imbed the random shocks component as far as we can rewrite (8) as :

$$Z_t=\alpha Z_{t-1} \tag{11}$$

We can suppose the sequence Z_t , Z_{t-1} ,.... as a sequence of random oscillations related by equation (9) so that we can affirm that there is a dynamic link which can model the random shocks in a time interval .Equation (9) suggests a markovian evolution of the random component so that we can suppose that we can write :

$$Z_{t+1}=\alpha Z_t= \alpha^2 Z_{t-1} = \ldots= \alpha^h Z_{t-h+1} \tag{12}$$

The insertion of the hypothesis on the stochastic components which are supposed represented by the difference between two consecutive determinations of the variable Y can give a deeper meaning to forecasting process.

As far as we limit our hypothesis to the expected value and to the long run behavior of the random component ,we can say something about the meaning and the risk of the futere value forecasting.

5 Forecasting and simulation.

We can try to exploit the previous results for the forecasting process and for the simulation process.

From (5) and (6) we state that two algebraic structures are related in order two obtain a forecasting with the horizon at t+4.

We try to generalize the previous results to whatever horizon in order to ensure the full generality of the mentioned approach.

Some further questions require an answer:

i. can we generalize equation (6) to any horizon ?
ii. can we exploit equation (6) for a simulation phase ?

Actually, we can try to ask the two questions considering that :

i. the actual generalization implies that we can use group theory for obtaining some insight into the forecasting process;
ii. equation (6) can be exploited as far as we choose a simulation strategy which can be either a deterministic or a stochastic one.

Group theory says something useful for our problem in so far as we can find an isomorphic model enjoying the same property of the mentioned model. Following the previous remarks we can consider the following mapping:

$$P_{h+k}(\alpha,\beta) * S_{h+k}(t) \Rightarrow c^{h+k} \exp[\, w(h+k)] \qquad (13)$$

The previous mapping implies that we can exploit the properties of the negative exponential function which can be considered the negative exponential distribution too, provided the coefficient c^{h+k} is a normalizing factor and the exponents are modified so that the mapping is preserved .
The previous mapping is allowed by the algebraic properties of both the polynomial operator coupled with the shift operator and the negative distribution operator.
We can exploit the right side of the mapping so that a stochastic random numbers generator can be built up which is able to simulate the future behavior of the variable Y,[6] and [7].
Actually the coefficients (c,w) can be suitably chosen so that the properties of the negative exponential distribution are assured.

6 Some operational implications.

The previous remarks allow some relevant features of the proposed approach.
We consider the negative exponential distribution as the source of the radom shocks related to the exponential smoothing model (5). This approach is not very common but following our opinion is able to justify the role of the random shocks in the exponential smoothing model.
Let us consider previous equations (5),(6),(7).As far as the homomorphic representation preserves the mentioned axiomatic properties, it is possible to generate future values both by the solution of the recursive equation (6) and by the negative exponential distribution provided we can estimate the parameters of both the models.
This result seems quite new in so far as the exploitation of the mapping (12) ensures the simulation of future scenarios as well as can be considered valid support for a forecasting process without passing through a cumbersome iteration of equation (6) starting from a given horizon.
We can put some questions about the role of the available data, that is the available time series. We mentioned in the previous pages that the parameters of the negative exponential distribution must be estimated and the available data can be used for estimating the parameters. In order to give an rough idea of the possible strategy we can consider that the exponent can be modified in order to fit the usual requirements of a negative exponential distribution. In other terms we must evaluate the expected value of the variable Y along the available interval of time so that the exponent can be reduced to the sum of two intervals of time h and k .In order to obtain this result we can consider

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

$w = Y/\Sigma Y_i$ as well as $c = \Sigma \exp [-w (h+k)]$.
The previous results allow the forecasting of future values of Y provided we fix in advance $h+k = M$.

7 Concluding remarks

The paper has the aim of proposing a new strategy for forecasting and simulating future values of a target variable Y provided we have at disposal a time series of a given length exploiting the homomorphic mapping between the class of the polynomial operators and the class of the exponential operators. Usually exponential smoothing process passes through the estimation of a parameter of a basic equation like (1) and the stochastic component plays a marginal role.
We show that we can introduce easily the random walk assumption on the stochastic component which can be related to the oscillations of the actual time series. Moreover a markov property seems the obvious consequence of the detection of the stochastic component. Finally, the algebraic properties of the operators identified in the iteration process which are preserved under an homomophic mapping on the set of the negative exponential functions as well as the negative exponential distributions, allows the solution of the forecasting process as well of the simulation process just by a random selection from the negative exponential distribution. The estimation process of the parameters of the negative exponential distribution can be routinely executed exploiting the actual time series whose length plays a marginal role in respect of the usual forecasting techniques.

Data Mining II, C.A. Brebbia & N.F.F. Ebecken (Editors)
© 2000 WIT Press, www.witpress.com, ISBN 1-85312-821-X

References

[1] Brown ,R.G. *Smoothing, Forecasting and Prediction of Discrete Time* Series ,Englewood Cliffs, N. J., Prentice Hall,1962.

[2] Brown R.G. *Statistical Forecasting for Inventory Control*; McGraw-Hill ,New York,1959.

[3] Mostow,G.D.,Sampson J.H. ,Meyer J.P. *Fundamental Structures of Algebra* ,McGraw-Hill, New York,1963.

[4] Arnold L. Dacuna-Castelle,D. *Random Dynamical Sytems*,Springer-Verlag,Berlin,19998.

[5] Tijas H.C. *Stochastic models,An Algorithmic Approach*. ,Wiley, Chichester,1994.

[6] Rubinstein,R.Y. *Simulation and the Monte Carlo Method*, Springer--Verlag, New York.,1981

[7] Bratley P.,Fox B.L., Schrage L.E. *A Guide to Simulation*, Springer-Verlag, New York,1983.