Chapter 18
Parameter Estimation of the Exponential Smoothing with Independent Regressors Using R

Ahmad Farid Osman

Abstract As an extension to the state space approach of the exponential smoothing methods, Osman and King recently introduced a new class of exponential smoothing methods by integrating regressors into the model. For the model to be utilized successfully, it requires a proper estimation procedure. The parameter estimation can be done through optimization using the "optim" function available in R statistical software. The objective of this paper is to discuss the effective use of the "optim" function in estimating parameters of the state space model of the exponential smoothing method augmented with regressors. The study started by considering several sets of optimization R codes to be supplied to the "optim" function. These codes use different set of initial values as the starting points for the optimization routine. The other difference between optimization codes is also in terms of restrictions imposed on the optimization routine. The second phase of the study was done by generating a number of simulated time series data with predetermined parameter values. The final phase of the study was then conducted by applying the optimization codes on all simulated series. By analyzing the performance of each of the optimization codes to accurately estimate parameters, a guideline or suggestion on how to effectively execute "optim" function in R with respect to the use of the new forecasting approach then outlined.

Keywords Optimization · Time-varying parameter · State-space model · Parameter space

A.F. Osman (✉)
Department of Applied Statistics, Faculty of Economics and Administration,
Universiti Malaya, Kuala Lumpur, Malaysia
e-mail: faridosman@um.edu.my

18.1   Introduction

The exponential smoothing method is an approach that used to produce future forecasts for a univariate time series. The method is believed to have been around since 1950s. The classical forms of exponential smoothing method are including simple exponential smoothing [1], linear [2] and Holt–Winter's additive and multiplicative [3]. Following the idea of nonlinear state-space framework, characterized by a single source of errors introduced by [4], the exponential smoothing method can now be expressed via state space representation. Hyndman et al. [5] thoroughly explain the state-space approach of the exponential smoothing method. A total of 30 possible models makes this approach to be very useful in producing forecasts for any univariate time series data. The disadvantage of this approach, however, it does not allow for integration of regressor(s) into the model. To mitigate this problem, Osman and King [6] introduced an extended version in which regressor(s) can be included into the model. By taking an example of a basic model that consist of one regressor, $z_{1,t}$, the method with time varying parameter of regressor can be expressed as follows:

$$y_t = \ell_{t-1} + b_{1,t-1}\Delta_{z_{1,t}} + \varepsilon_t \tag{18.1a}$$

$$\ell_t = \ell_{t-1} + b_{1,t-1}\Delta_{z_{1,t}} + \alpha\varepsilon_t \tag{18.1b}$$

$$b_{1,t} = \begin{cases} b_{1,t-1} + \frac{\beta_1\left(\varepsilon_{1,t-1}^+ + \varepsilon_t\right)}{\Delta_{z_{1,t}}^*}, & \text{if } \left|\Delta_{z_{1,t}}^*\right| \geq L_{b_1} \\ b_{1,t-1}, & \text{if } \left|\Delta_{z_{1,t}}^*\right| < L_{b_1} \end{cases} \tag{18.1c}$$

$$\varepsilon_{1,t}^+ = \begin{cases} 0, & \text{if } \left|\Delta_{z_{1,t}}^*\right| \geq L_{b_1} \\ \varepsilon_{1,t-1}^+ + \varepsilon_t, & \text{if } \left|\Delta_{z_{1,t}}^*\right| < L_{b_1}, \end{cases} \tag{18.1d}$$

where

$$\Delta_{z_{1,t}}^* = z_{1,t} - z_{1,t-1}^* \quad \text{and} \quad z_{1,t}^* = \begin{cases} z_{1,t}, & \text{if } \left|\Delta_{z_{1,t}}^*\right| \geq L_{b_1} \\ z_{1,t-1}^*, & \text{if } \left|\Delta_{z_{1,t}}^*\right| < L_{b_1} \end{cases}. \tag{18.2}$$

ℓ_t in Eq. (18.1) denotes the level of the series, $b_{1,t}$ represents a regressor parameter, ε_t denotes errors, while α and β_1 represent the smoothing parameters. Both $\varepsilon_{1,t}^+$ and $z_{1,t}^*$ are representing dummy errors and dummy regressor, respectively. On the other hand, L_{b_1} represents the lower boundary for a "switching procedure." Detailed explanation on this model is given in Osman and King [6].

Given that this is a new developed method, there is no available statistical software that can be used to execute parameter estimation for the method at this

moment. In general, the objective of this study is to determine how best of executing "optim" function in R statistical software to estimate parameters of the new models.

18.2 Research Methodology

There are two main issues involved in estimating parameters of Eq. (18.1). First, what constraints need to be imposed on the smoothing constants, α and β_1. Second, what values of starting points (for the two smoothing constants) should be used in optimization routine. In order to find answer for the first issue, two approaches were considered that provide different parameter space for the two smoothing constants. The first approach is to impose parameter space according to the classical boundary of the smoothing constants in exponential smoothing method, that is between 0 and 1 for α and between 0 and α for β_1 as explained by [6] and Osman and King [3]. The second approach is to impose restrictions on eigenvalues of the characteristic equation of the model to be within the unit circle as explained by Osman and King [7]. The second approach is known as forecastability concept which gives wider parameter space for the smoothing constants as compared to the first approach. With regard to the second issue, different sets of starting points for the smoothing constants were considered in this analysis that is either 0.01, 0.1, or 0.5 for each of the smoothing constants.

The analysis which only involves one model as described by Eq. (18.1) was started by producing two optimization R codes for parameter estimation using two different sets of restrictions as explained above. The next step then followed to produce six simulated series with different predetermined values of parameters as listed in Table 18.1. For all cases of simulation, initial level and growth term were set to be $\ell_0 = 500$ and $b_{1,0} = 0.5$. By considering a situation where the "switching procedure" is not required, all simulated series were generated based on the following equations where the error term, ε_t is a normally distributed generated series.

$$y_t = \ell_{t-1} + b_{1,t-1}\Delta_{z_{1,t}} + \varepsilon_t \tag{18.3a}$$

$$\ell_t = \ell_{t-1} + b_{1,t-1}\Delta_{z_{1,t}} + \alpha\varepsilon_t \tag{18.3b}$$

$$b_{1,t} = b_{1,t-1} + \beta_1\varepsilon_t/\Delta_{z_{1,t}} \tag{18.3c}$$

Table 18.1 Predetermined parameter values for six simulated series

	α	β_1		α	β_1
Simulation 1	0.10	0.01	Simulation 4	0.50	0.10
Simulation 2	0.10	0.10	Simulation 5	0.90	0.01
Simulation 3	0.50	0.01	Simulation 6	0.90	0.10

Table 18.2 Starting points used in optimization routine

α	β_1	Possible combination of starting points considered in optimization, (α, β_1)
0.01	0.01	(0.01, 0.01)
0.10	0.01	(0.10, 0.01)
0.10	0.10	(0.10, 0.10)
0.50	0.01	(0.50, 0.01)
0.50	0.10	(0.50, 0.10)
0.50	0.50	(0.50, 0.50)
0.01, 0.10	0.01, 0.10	(0.01, 0.01), (0.01, 0.10), (0.10, 0.01), (0.10, 0.10)
0.01, 0.10, 0.50	0.01, 0.10, 0.50	(0.01, 0.01), (0.01, 0.10), (0.01, 0.50), (0.10, 0.01), (0.10, 0.10), (0.10, 0.50), (0.50, 0.01), (0.50, 0.10), (0.50, 0.50)

The third stage of analysis was done by applying optimization R codes for parameter estimation on all simulated series. A total of eight sets of starting points for smoothing constants were used with six of them that are simply combination between single starting point for both α and β_1. For other two sets of starting points, either two or three different values of starting points were used for each smoothing constant. All possible combinations of starting points are then considered in the optimization routine. Table 18.2 depicts all starting points used in estimation process. Note that, even though multiple starting points were used, only the one that gives the smallest value of the objective function was examined.

The estimation process was performed iteratively by minimizing the objective function $\left(n\log\left(\sum_{t=1}^{n} \varepsilon_t^2 \right) \right)$ of the optimization code. This optimization procedure was conducted using "optim" function with the use of Nelder–Mead algorithm. The final step of analysis was then to evaluate the estimation results by comparing the predetermined values and the estimated values of all parameters.

18.3   Estimation Results

Estimation results are given in Tables 18.3, 18.4, 18.5, 18.6, 18.7, and 18.8. It can be seen that in most cases, small values of starting points ($\alpha = 0.01$, $\beta_1 = 0.01$ and $\alpha = 0.1$, $\beta_1 = 0.01$) managed to produce the smallest value of the objective function. This finding, however, is not true for simulation 5 and also not true for simulations 3 and 4 in which restrictions on parameter space were imposed according to the forecastability concept. Another finding is that, when the actual parameters (predetermined values) are small, the use of large starting points has led to the failure of the optimization routine to produce estimation that close to the actual values as evident by the results of analysis in the case of simulation 1 and 2. As mentioned earlier, estimation based on the forecastability concept provides

Table 18.3 Estimation results for simulated series 1

Classical boundary						
Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	490.5707	0.5042	0.0610	0.0080	**882.0783**
0.10	0.01	490.5887	0.5042	0.0609	0.0080	**882.0783**
0.10	0.10	491.3279	0.5037	0.0607	0.0079	882.1475
0.50	0.01	489.1021	0.5033	**0.4089**	0.0135	894.1491
0.50	0.10	492.9763	0.5024	0.0660	0.0079	882.7981
0.50	0.50	490.5314	0.5044	0.0596	0.0080	882.0814
0.01, 0.10		490.5887	0.5042	0.0609	0.0080	**882.0783**
0.01, 0.10, 0.50		490.5887	0.5042	0.0609	0.0080	**882.0783**
Forecastability concept						
Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	489.9522	0.5060	**2.18E-06**	**0.0049**	886.3208
0.10	0.01	490.5844	0.5042	0.0608	0.0080	**882.0783**
0.10	0.10	490.6940	0.5044	0.0601	0.0082	882.1154
0.50	0.01	493.5977	0.5023	0.0696	0.0084	883.2511
0.50	0.10	490.6622	0.5042	0.0602	0.0080	882.0793
0.50	0.50	489.4188	0.5053	0.0590	0.0078	882.3468
0.01, 0.10		490.5844	0.5042	0.0608	0.0080	**882.0783**
0.01, 0.10, 0.50		490.5844	0.5042	0.0608	0.0080	**882.0783**

Table 18.4 Estimation results for simulated series 2

Classical boundary						
Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	499.6713	0.5437	0.0965	0.0965	888.2322
0.10	0.01	496.6999	0.5679	0.0987	0.0986	**885.1003**
0.10	0.10	494.0989	0.6097	0.0986	0.0986	887.5330
0.50	0.01	497.1064	0.5793	0.0968	0.0968	887.3669
0.50	0.10	498.8807	0.5607	0.0961	0.0961	887.9756
0.50	0.50	496.0680	0.5655	**0.2274**	0.0970	899.7527
0.01, 0.10		496.6999	0.5679	0.0987	0.0986	**885.1003**
0.01, 0.10, 0.50		496.6999	0.5679	0.0987	0.0986	**885.1003**
Forecastability concept						
Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	496.8073	0.5991	0.0442	0.0978	**880.7533**
0.10	0.01	496.0696	0.5654	**0.2279**	0.0970	899.7527
0.10	0.10	493.4104	0.6271	0.0551	0.1029	882.9565
0.50	0.01	496.0702	0.5655	**0.2274**	0.0970	899.7527
0.50	0.10	492.9681	0.5885	0.1120	0.1032	891.7968
0.50	0.50	496.6072	0.5614	**0.2319**	0.0970	899.7886
0.01, 0.10		496.8073	0.5991	0.0442	0.0978	**880.7533**
0.01, 0.10, 0.50		496.8073	0.5991	0.0442	0.0978	**880.7533**

Table 18.5 Estimation results for simulated series 3

Classical boundary

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	485.8148	0.5060	0.6596	0.0136	**883.3676**
0.10	0.01	478.3717	0.5063	0.6515	0.0133	884.3208
0.10	0.10	485.8522	0.5059	0.6595	0.0136	**883.3676**
0.50	0.01	485.8272	0.5060	0.6591	0.0136	**883.3676**
0.50	0.10	485.8050	0.5059	0.6593	0.0136	**883.3676**
0.50	0.50	478.6677	0.5062	0.6550	0.0133	884.2457
0.01, 0.10		485.8148	0.5060	0.6596	0.0136	**883.3676**
0.01, 0.10, 0.50		485.8148	0.5060	0.6596	0.0136	**883.3676**

Forecastability concept

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	485.0704	0.5061	0.6701	0.0131	883.4157
0.10	0.01	476.6314	0.5065	0.6436	0.0132	884.8260
0.10	0.10	478.0573	0.5062	0.6586	0.0133	884.4020
0.50	0.01	485.8229	0.5060	0.6596	0.0136	**883.3676**
0.50	0.10	485.8287	0.5060	0.6599	0.0136	**883.3676**
0.50	0.50	479.5304	0.5062	0.6512	0.0133	884.0483
0.01, 0.10		485.8167	0.5060	0.6597	0.0136	**883.3676**
0.01, 0.10, 0.50		485.8229	0.5060	0.6596	0.0136	**883.3676**

Table 18.6 Estimation results for simulated series 4

Classical boundary

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	491.8341	0.5654	0.5788	0.1019	**883.1337**
0.10	0.01	491.8140	0.5654	0.5787	0.1019	883.1338
0.10	0.10	491.8329	0.5655	0.5788	0.1019	883.1338
0.50	0.01	491.9181	0.5669	0.5814	0.1016	883.1428
0.50	0.10	491.8227	0.5655	0.5788	0.1020	883.1338
0.50	0.50	491.8329	0.5654	0.5787	0.1019	**883.1337**
0.01, 0.10		491.8341	0.5654	0.5788	0.1019	**883.1337**
0.01, 0.10, 0.50		491.8341	0.5654	0.5788	0.1019	**883.1337**

Forecastability concept

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	491.8447	0.5654	0.5787	0.1020	883.1338
0.10	0.01	491.8128	0.5654	0.5787	0.1019	883.1338
0.10	0.10	491.8210	0.5655	0.5787	0.1019	883.1338
0.50	0.01	491.8337	0.5653	0.5788	0.1019	**883.1337**
0.50	0.10	491.8399	0.5653	0.5786	0.1019	883.1338
0.50	0.50	491.8136	0.5655	0.5788	0.1019	883.1338
0.01, 0.10		491.8128	0.5654	0.5787	0.1019	883.1338
0.01, 0.10, 0.50		491.8337	0.5653	0.5788	0.1019	**883.1337**

Table 18.7 Estimation results for simulated series 5

Classical boundary

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	466.0406	0.4934	1.0000	0.0135	889.1787
0.10	0.01	484.8486	0.4929	0.9993	0.0169	883.9131
0.10	0.10	466.5630	0.4942	1.0000	0.0140	888.8960
0.50	0.01	473.3510	0.4972	1.0000	0.0137	885.8410
0.50	0.10	486.2458	0.4965	1.0000	0.0143	**883.3803**
0.50	0.50	466.0930	0.4936	1.0000	0.0137	889.1486
0.01, 0.10		484.8486	0.4929	0.9993	0.0169	883.9131
0.01, 0.10, 0.50		486.2458	0.4965	1.0000	0.0143	**883.3803**

Forecastability concept

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	487.7907	0.4972	**1.1450**	0.0137	**881.5762**
0.10	0.01	487.7899	0.4972	**1.1447**	0.0138	**881.5762**
0.10	0.10	487.7884	0.4973	**1.1446**	0.0137	**881.5762**
0.50	0.01	489.8414	0.5011	**1.1408**	0.0150	881.7747
0.50	0.10	487.7656	0.4972	**1.1446**	0.0137	**881.5762**
0.50	0.50	487.8659	0.4972	**1.1493**	0.0137	881.5781
0.01, 0.10		487.7907	0.4972	**1.1450**	0.0137	**881.5762**
0.01, 0.10, 0.50		487.7656	0.4972	**1.1446**	0.0137	**881.5762**

Table 18.8 Estimation results for simulated series 6

Classical boundary

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	488.7248	0.5210	0.9672	0.1045	884.9660
0.10	0.01	488.6802	0.5210	0.9670	0.1046	**884.9659**
0.10	0.10	488.6616	0.5211	0.9670	0.1046	884.9659
0.50	0.01	488.6761	0.5210	0.9668	0.1046	884.9660
0.50	0.10	488.6834	0.5210	0.9671	0.1046	**884.9659**
0.50	0.50	488.6856	0.5211	0.9671	0.1046	**884.9659**
0.01, 0.10		488.6802	0.5210	0.9670	0.1046	**884.9659**
0.01, 0.10, 0.50		488.6834	0.5210	0.9671	0.1046	**884.9659**

Forecastability concept

Starting points		Estimated coefficients				Objective function
α	β_1	ℓ_0	$b_{1,0}$	α	β_1	
0.01	0.01	488.6967	0.5210	0.9671	0.1046	**884.9659**
0.10	0.01	488.6582	0.5210	0.9670	0.1046	**884.9659**
0.10	0.10	488.6894	0.5211	0.9671	0.1046	**884.9659**
0.50	0.01	488.6784	0.5210	0.9671	0.1046	**884.9659**
0.50	0.10	488.6732	0.5211	0.9671	0.1046	**884.9659**
0.50	0.50	488.6926	0.5211	0.9668	0.1045	884.9660
0.01, 0.10		488.6967	0.5210	0.9671	0.1046	**884.9659**
0.01, 0.10, 0.50		488.6784	0.5210	0.9671	0.1046	**884.9659**

wider space for the smoothing constants. If we look at the estimation result of simulation 5, the estimated values of α when estimated based on the forecastabilty concept are much higher than the predetermined value. In this particular case, estimation performed based on the classical boundary method produced more accurate estimated value for α.

18.4   Conclusion

As discussed in the previous section, the use of small values of starting points for smoothing constants in optimization routine is advised as compared to large values of starting points. However, taking into account the better performance of large starting points in some cases, it is encouraged to use multiple starting points that is also include small values of starting points. It is also can be concluded that the use of forecastability concept in imposing restrictions on parameter space does not necessarily will produce better estimate than the use of classical boundary concept. Since it provides wider parameter space for the smoothing constant, it allows for greater chance of estimation error as well.

The use of classical boundary concept in imposing restrictions to the parameter space of the smoothing constant is actually sufficient if we want to update the levels, ℓ_t and the regressor parameter, $b_{1,t}$ based on the weighted average methodology, the concept used in the classical form of the exponential smoothing methods. This is because restricting α to be between 0 and 1 and β_1 to be between 0 and α in the state-space representation is equivalent to restrict both α and β_1 to be between 0 and 1 in the classical form of the exponential smoothing methods.

References

1. Brown, R.G.: Statistical Forecasting for Inventory Control. McGraw-Hill, New York (1959)
2. Holt, C.C.: Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages. Office of Naval Research Memorandum, vol 52 (1957)
3. Winters, P.R.: Forecasting sales by exponentially weighted moving averages. Manage. Sci. **6**, 324–342 (1960)
4. Ord, J.K., Koehler, A.B., Synder, R.D.: Estimation and prediction for a class of dynamic nonlinear statistical models. J. Am. Statist. Assoc. **92**, 1621–1629 (1997)
5. Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D.: Forecasting with Exponential Smoothing: The State Space Approach. Springer Verlag, Berlin (2008)
6. Osman, A.F., King, M.L.: A New Approach to Forecasting based on Exponential Smoothing with Independent Regressors. Working paper series, Monash University (2015)
7. Osman, A.F., King, M.L.: Exponential smoothing with regressors: estimation and initialization. Model Assist. Stat. Appl. **10**(3), 253–263 (2015)