Exponential Smoothing Based on Ranks

J. Nyblom[1]

[1] University of Joensuu, P.O.Box 111, FIN-80101 Joensuu, Finland

Keywords: Efficiency, Local level model, Outliers, Prediction, Robustness.

1 Exponential smoothing

1.1 Definition

Exponential smoothing is widely used forecasting method first suggested by C. C. Holt in the late 1950's. An elementary introduction is provided by Chatfield (2004).

Let $(y_1, y_2, \ldots, y_N) = \mathbf{y}^N$ be a non-seasonal time series with no systematic trend, then we may try to forecast y_{N+1} by exponentially smoothed past values as

$$\hat{y}_{N+1} = ay_N + a(1-a)y_{N-1} + a(1-a)^2 y_{N-2} + \cdots. \tag{1}$$

The forecast satisfies a recurrence formula

$$\hat{y}_{N+1} = ay_N + (1-a)\hat{y}_N \tag{2}$$

which is applicable also to a finite sequence with an appropriate starting value for $\hat{y}_2$. Customarily we choose $\hat{y}_2 = y_1$.

1.2 Optimality

Assume the local level model

$$y_t = \mu_t + \varepsilon_t, \tag{3}$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad t = 1, 2, \ldots \tag{4}$$

where the random errors ε_t and η_t are independent normal variables with mean zero and variances σ_ε^2 and σ_η^2, respectively. Denote the signal-to-noise ratio as $q = \sigma_\eta^2/\sigma_\varepsilon^2$. Then the best forecasts for y_{N+1} and μ_{N+1} are equal to their conditional expectations given the past $\mathbf{y}^N$. We easily find that they coincide. If N is not small $E(y_{N+1} \,|\, \mathbf{y}^N)$ is well approximated by the recursions described above, if we take $a = (-q + \sqrt{q^2 + 4q})/2$, see Harvey (1993, p. 127).

On the other hand if the generating process is likely to produce outlying observations, the procedure may be far from optimal. The next section defines a rank based method that is preferable in such situations. We use the model (3)–(4) as a test bed allowing outliers among ε_t's but retaining η_t's normal. Since, by definition, a robust method will produce inaccurate predictions for outlying observations, the performance of competing methods is measured by the accuracy of predicting the more stable latent process μ_t.

2 Exponential smoothing using ranks

2.1 Definition

As a first step we apply exponential smoothing on ranks. Let $r_{1:N}, \ldots, r_{N:N}$ be the mid-ranks of $y_1, y_2, \ldots, y_N$, and let $\hat{r}_{N,N-1}$ be the forecast of $r_{N:N}$ based on the ranks up to time $N-1$. Then, by analogy with (2), we update the rank forecast as

$$\hat{r}_{N+1,N} = ar_{N:N} + (1-a)\hat{r}_{N,N-1}. \tag{5}$$

Note that if $\hat{r}_{N,N-1} \le N-1$ then $\hat{r}_{N+1,N} < N$. Thus, by induction, $\hat{r}_{2,1} = 1$ generates a series with $\hat{r}_{N+1,N} < N$ for $N = 2, 3, \ldots$.

Because we are interested in forecasting the process itself not its ranks, we continue by writing $\hat{r}_{N+1,N} = k + \theta$ where $0 \le \theta < 1$ with $k < N$. Finally, the forecast is defined by

$$\tilde{y}_{N+1} = (1-\theta) y_{k:N} + \theta y_{k+1:N}, \tag{6}$$

where $y_{1:N} \le \cdots \le y_{N:N}$ are the ordered values of $y_1, y_2, \ldots, y_N$.

2.2 Updating formulas

In addition to keeping records on the values $\hat{r}_{2,1}, \hat{r}_{3,2}, \ldots$ we have to update the ordered series $y_{1:N} \le \cdots \le y_{N:N}$, $N = 1, 2, \ldots$. When the new value y_N arrives, its position in the updated ordered vector is given by the low rank

$$\underline{r}_{N:N} = 1 + \sum_{t=1}^{N-1} I(y_t < y_N), \tag{7}$$

where $I(\cdot)$ is the indicator function. The mid-rank is given by

$$r_{N:N} = 1 + \sum_{t=1}^{N-1} \left[I(y_t < y_N) + \frac{1}{2} I(y_t = y_N) \right], \tag{8}$$

3 Efficiency

A limited simulation experiment has been made for a comparison between the new method and the standard exponential smoothing. The local level model (3)–(4) is employed, and the performance is measured by the MSE and MAD criterions

$$\sum_{t=2}^{N} (\hat{y}_t - \mu_t)^2, \quad \sum_{t=2}^{N} (\tilde{y}_t - \mu_t)^2, \quad \sum_{t=2}^{N} |\hat{y}_t - \mu_t|, \quad \sum_{t=2}^{N} |\tilde{y}_t - \mu_t| \tag{9}$$

for the ordinary and rank-based exponential smoothing, respectively. The coefficient a in each replicate is optimized over the grid $a = 0.1, \ldots, 0.9$. The optimization is carried out using the observations themselves for the ordinary exponential smoothing and using ranks for the new rank version. Then ratios of MSE's and MAD's are reported in the TABLE 1. The figures can be interpreted as efficiency of the new method. We find that under normality the efficiency is above 80%. In the presence of outliers the new method is more efficient except when $q = 10$, i.e. when the random walk is relatively volatile. The outlying ε_t's are randomly positioned and set to $\pm 10\sigma_\varepsilon$. The rest of the errors are from $N(0, \sigma_\varepsilon^2)$.

TABLE 1. Efficiencies of the rank EWMA; the length of the series is $N = 50$.

	Normal		Five outliers		One outlier	
q	MSE	MAD	MSE	MAD	MSE	MAD
0.1	0.821	0.908	1.383	1.369	1.111	1.066
1.0	0.873	0.935	1.343	1.208	1.091	1.023
10	0.990	0.995	0.943	0.986	0.960	0.987

References

C. Chatfield (2004). *The Analysis of Time Series — An Introduction*, 6th Edition. Chapman and Hall, London.

A. Harvey (1993). *Time Series Models*, 2nd Edition. Harvester Wheatsheaf, New York.