Anomaly Detection

Preetam Jinka

April 2, 2015

1 Introduction

The Holt-Winters method is the method of triple exponential smoothing. The exponential moving average filter applied three times accounts for level, trend, and seasonality for a given time series. The single exponentially weighted moving average is defined as

$$s_t = \alpha x_t + (1 - \alpha) s_{t-1}. \tag{1}$$

We can introduce another smoothing parameter, β, to account for a linear trend.

$$\begin{aligned} s_t &= \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta) b_{t-1} \end{aligned} \tag{2}$$

Finally, to account for a multiplicative seasonality, we introduce another smoothing parameter, γ.

$$\begin{aligned} s_t &= \alpha \frac{x_t}{g_{t-L}} + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta) b_{t-1} \\ g_t &= \gamma \frac{x_t}{s_t} + (1 - \gamma) g_{t-L} \end{aligned} \tag{3}$$

For our purposes, we will not consider seasonality.

1.1 "Anomaly detection"

We define "anomalies" in time series as statistically unlikely values given a limited number of previous terms. The process of detecting these anomalies occurs every time a new value is observed.

2 One step-ahead prediction interval

In the context of anomaly detection, we are interested in detecting points that fall outside previously computed prediction intervals. For simplicity, we consider an AR(1) time series and compared the AR prediction interval with the Holt-Winters approach.

2.1 Simulations

The AR(1) process is given by

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{4}$$

We first generate N terms of an AR(1) process using the `arima.sim` function in R. The smoothed time series, s_t, is then computed iteratively using either (1) or (2).

The square-error time series is defined as

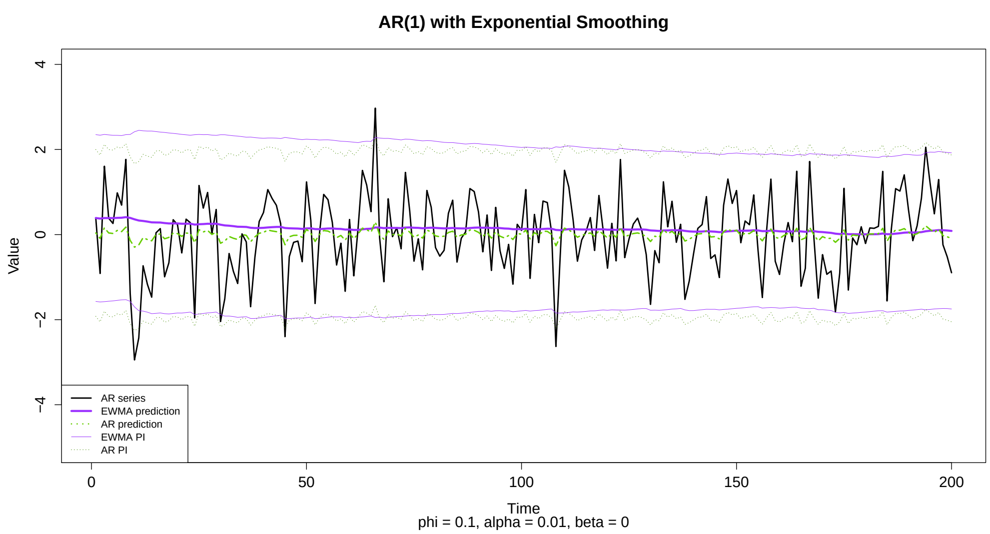
$$e_t = (1-\alpha)e_{t-1} + \alpha(s_t - x_t)^2 \tag{5}$$

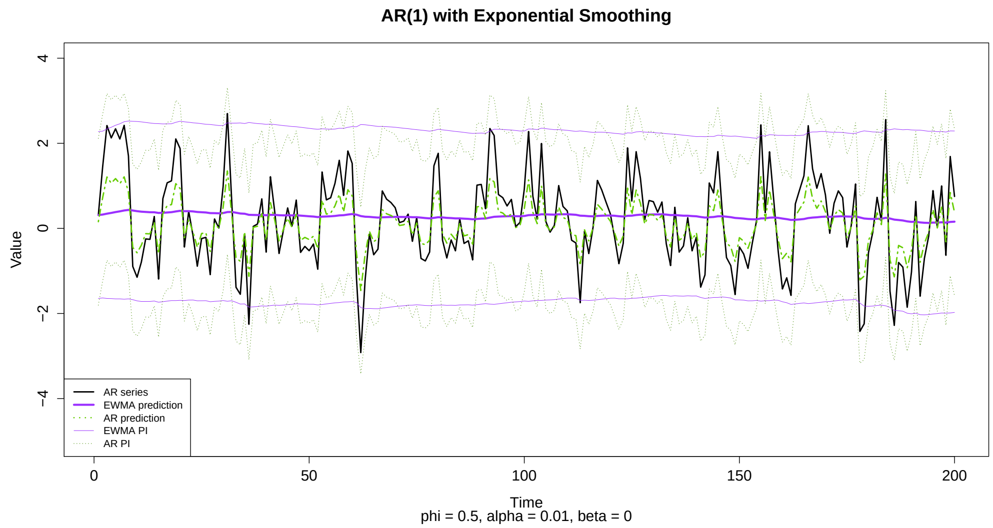
e_t serves as an exponentially-weighted moving MSE.

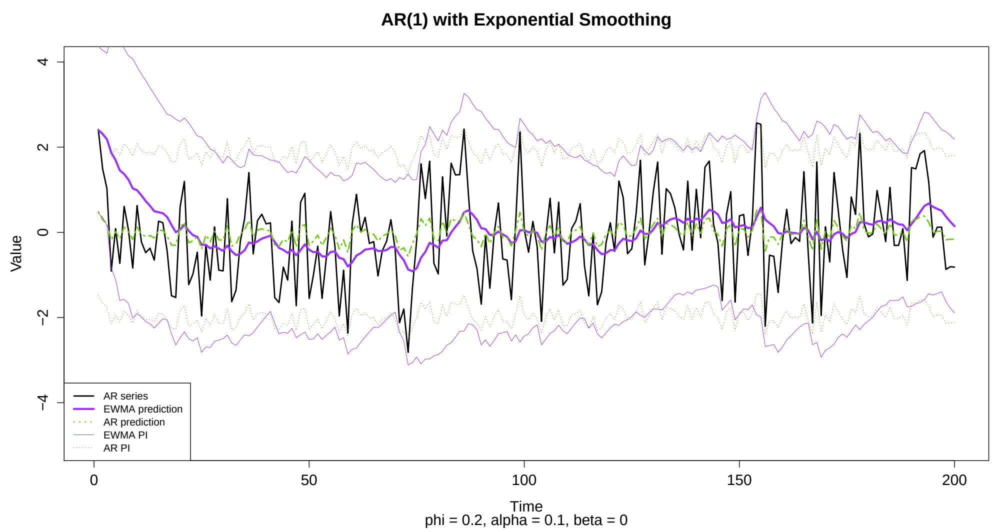
The prediction intervals are computing using

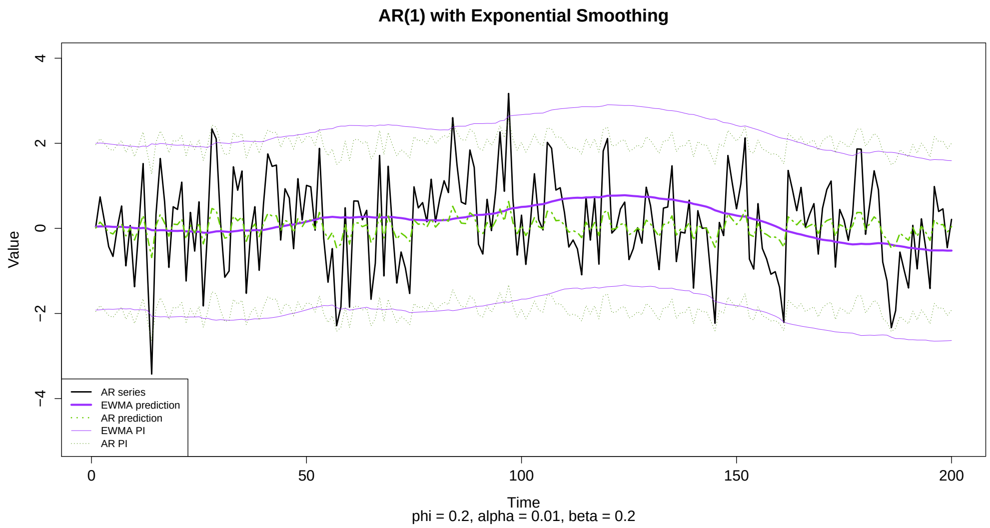
$$\text{Prediction interval} = s_t \pm 1.96\sqrt{e_t} \tag{6}$$

The following are plots of a simulated AR(1) with single and double exponential smoothing. Each plot also contains the AR(1) and EWMA predicted values, and the respective prediction intervals.




AR(1) with Exponential Smoothing
Value
Time
phi = 0.5, alpha = 0.01, beta = 0
AR series
EWMA prediction
AR prediction
EWMA PI
AR PI


AR(1) with Exponential Smoothing
Value
Time
phi = 0.2, alpha = 0.1, beta = 0
AR series
EWMA prediction
AR prediction
EWMA PI
AR PI



2.2 Coverage probabilities

The coverage probabilities were tested against the AR(1) process described in the previous section. In order to calculate the coverage probability, we calculate the number of times a newly observed value falls outside of the predicted interval for that value. First, we simulate an AR(1) process as before. We run the simulation and stop at least one step before the end. Finally, we compare the next simulated value against the prediction interval of the smoothed series. We increment a counter if the value falls outside of the prediction interval. This process is repeated for at least 5000 iterations.

As the table shows, the choices of α and β greatly affect the overall coverage probabilities. In practice, many combinations of these parameters should be checked in order to select optimal values.

Iterations	α	β	Expected	Actual	Absolute difference
5000	0	0	0.05	0.165	0.115
5000	0.05	0	0.05	0.0568	0.0068
5000	0.05	0.05	0.05	0.0528	0.0028
10000	0.25	0.05	0.05	0.0876	0.0376
10000	0.25	0.25	0.05	0.0857	0.0357

3 Smoothed MA(1)

Let x_t be MA(1).

$$
\begin{aligned}
\mathrm{E}(x_t) &= 0 \\
\mathrm{Var}(x_t) &= (\theta^2 + 1)\sigma^2 \\
\mathrm{Cov}(x_t, x_{t+h}) &= \begin{cases} \sigma^2(1+\theta^2), & \text{if } h = 0. \\ \sigma^2\theta, & \text{if } h = 1. \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{7}
$$

We define s_t as x_t with single exponential smoothing with parameter α, i.e.

$$
s_t = \alpha x_t + (1-\alpha)s_{t-1}. \tag{8}
$$

3.1 Iterative form of s_t

Using (8), we calculate the first few terms of s_t.

$$
\begin{aligned}
s_1 &= (1-\alpha)s_0 + \alpha x_1 \\
s_2 &= [(1-\alpha)s_0 + \alpha x_1](1-\alpha) + \alpha x_2 \\
&= (1-\alpha)^2 s_0 + \alpha(1-\alpha)x_1 + \alpha x_2 \\
s_3 &= \left[(1-\alpha)^2 s_0 + \alpha(1-\alpha)x_1 + \alpha x_2\right](1-\alpha) + \alpha x_3 \\
&= (1-\alpha)^3 s_0 + \alpha(1-\alpha)^2 x_1 + \alpha(1-\alpha)x_2 + \alpha x_3
\end{aligned} \tag{9}
$$

This can be rewritten in iterative form as

$$
s_t = (1-\alpha)^t s_0 + \alpha \sum_{i=1}^{t} (1-\alpha)^{t-i} x_i. \tag{10}
$$

3.2 Variance of s_t

We will derive the closed-form variance of s_t. Note that s_0 is assumed to be constant in the following equation.

$$
\begin{aligned}
\mathrm{Var}(s_t) &= \mathrm{Var}\left((1-\alpha)^t s_0 + \alpha \sum_{i=1}^{t} (1-\alpha)^{t-i} x_i\right) \\
&= \mathrm{Var}\left(\alpha \sum_{i=1}^{t} (1-\alpha)^{t-i} x_i\right)
\end{aligned} \tag{11}
$$

Recall the following identity

$$
\mathrm{Var}\left(\sum_{i=1}^{t} x_i\right) = \sum_{i=1}^{t} \mathrm{Var}(x_i) + \sum_{i \neq j} \mathrm{Cov}(x_i, x_j) \tag{12}
$$

Using (12), we can rewrite (11) as

$$
\begin{aligned}
\mathrm{Var}(s_t) &= \mathrm{Var}\left(\alpha \sum_{i=1}^{t} (1-\alpha)^{t-i} x_i\right) \\
&= \sum_{i=1}^{t} \mathrm{Var}\left(\alpha(1-\alpha)^{t-i} x_i\right) + \sum_{i \neq j} \mathrm{Cov}\left[\alpha(1-\alpha)^{t-i}, \alpha(1-\alpha)^{t-j}\right]
\end{aligned} \tag{13}
$$

We compute the two terms independently starting with the left.

$$
\sum_{i=1}^{t} \mathrm{Var}\left(\alpha(1-\alpha)^{t-i} x_i\right) = \sum_{i=1}^{t} \alpha^2 (1-\alpha)^{2t-2i} \mathrm{Var}(x_i)
$$

We can substitute the variance of x_i using (7).

$$
= \sigma^2(\theta^2 + 1) \sum_{i=1}^{t} \alpha^2 (1-\alpha)^{2t-2i} \tag{14}
$$

For the right term, we arrive at the following

$$
\sum_{i \neq j} \mathrm{Cov}\left[\alpha(1-\alpha)^{t-i}, \alpha(1-\alpha)^{t-j}\right] = \sum_{i \neq j} \alpha^2 (1-\alpha)^{2t-i-j} \mathrm{Cov}(x_i, x_j) \tag{15}
$$

Observe that using (7), we see that the only terms that contribute are those with i and j differing by 1. Due to symmetry, each of those terms contributes twice. Therefore, (15) can be rewritten as

$$
\sum_{i \neq j} \mathrm{Cov}\left[\alpha(1-\alpha)^{t-i}, \alpha(1-\alpha)^{t-j}\right] = 2\sigma^2\theta \sum_{i=1}^{t-1} \alpha^2 (1-\alpha)^{2t-(2i+1)}. \tag{16}
$$

Combining (14) and (16), we get

$$
\begin{aligned}
\mathrm{Var}(s_t) &= \mathrm{Var}\left(\alpha \sum_{i=1}^{t} (1-\alpha)^{t-i} x_i\right) \\
&= \sigma^2(\theta^2 + 1) \sum_{i=1}^{t} \alpha^2 (1-\alpha)^{2t-2i} + 2\sigma^2\theta \sum_{i=1}^{t-1} \alpha^2 (1-\alpha)^{2t-(2i+1)}
\end{aligned} \tag{17}
$$

□

3.3 Empirical results

The previously discussed simulation was modified to use a simulated MA(1) using `arima.sim`. For n iterations, the t-th value of the smoothed time series

s_t was collected. The variance of these collected values was compared to the closed-form variance derived in the previous section.

n	θ	σ	t	α	Iterative variance	Closed-form variance	Absolute difference
20000	0.1	1	1	1	1.006463	1.010000	0.00354
20000	0.1	1	1	0.5	0.2543537	0.2525000	0.00185
20000	0.1	1	10	0.5	0.3648053	0.3699996	0.00519
20000	0.1	1	10	0.1	0.05575114	0.05474687	0.001
20000	0.1	1	20	0.5	0.3649438	0.3700000	0.00506
20000	0.1	1	50	0.5	0.3778751	0.3700000	0.00788

References

[1] Prajakta S. Kalekar, *Time series Forecasting using Holt-Winters Exponential Smoothing*. Kanwal Rekhi School of Information Technology, 2004.