Holt-Winters methods

Marco Fattore

1 Introduction

As we have discussed in previous lessons, the deterministic approach to time series forecasting is too rigid to be useful in many practical situations. Also Exponential Smoothing does not solve the problem, being useful mainly for non-trended time series. To overcome these issues, still sticking to a heuristic, non-inferential approach, we now introduce the *Holt-Winters* (HW) method, which is a system of recursive equations which defines a model for the dynamic structure of a non-stationary time series, providing both a way to decompose the series into *hidden* or *unobserved components*, namely *trend* (or *level*) and *seasonal*, and to forecast it. The HW method can be seen as a way to perform *local regression*, interpolating linearly the time series, but adapting the coefficients of the linear function, over time. The link between past and future, imposed by the recursive HW relations, constraints the dynamics of these coefficients, allowing the computation of the components.

The HW method provides recursions for both trended and trended&seasonal time series; the latter model encompasses the former, but for didactic reasons, we introduce them separately, to clarify the ideas behind the approach.

2 Formal development

Let $y_1, \ldots, y_T$ be a time series; at each time t, the time series is implicitly assumed to be decomposed as the sum of the *level* ℓ_t, the seasonal component S_t and an error term e_t:

$$y_t = \ell_t + S_t + e_t. \tag{1}$$

However, as we see below, the error term is not explicitly inserted into the HW recursive equations and is not modeled in inferential terms. Above, we have introduced it, just to clarify that it is "conceptually" present, although not formally considered.

2.1 Non-seasonal Holt-Winters method

Let us begin with the trended, non-seasonal HW (THW) recursion, which is defined as follows:

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \quad (2)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1} \quad (3)$$

$$\hat{y}_{T+h|\le T} = \ell_T + h b_T \quad (4)$$

where $0 < \alpha, \beta < 1$ and h is a natural number.

The first equation computes the level at time t as a weighted average of the last observation y_t and the level we would forecast, based on the level at time $t-1$ (i.e., ℓ_{t-1}) and the *slope* at time $t-1$ (i.e., b_{t-1}). The slope expresses the linear increment of the level, over a unit of time; in a sense, it represents the "discrete derivative" of the level. So, the basic idea is that the level locally moves along a straight line, identified by the slope, but due to the effect of (not explicitly modeled) shocks, i.e. of the implicit error terms, it deviates from such a straight line. This is revealed by y_t being different from $\ell_{t-1} + b_{t-1}$. As a consequence, the level at time t must be corrected taking into account such a deviation. However, since we do not want the implicit error to be totally incorporated into the new value of the level (this would mean reproducing the fluctuations of the time series and not extracting a smoother hidden component), we average y_t and $\ell_{t-1} + b_{t-1}$, with a smoothing parameter α, which determines the relative importance of the last observation and of the past of the time series (which is incorporated into and summarized by the values of ℓ_{t-1} and b_{t-1}).

The second equation states a recursion for the slope term, which evolves and adapts over time, similarly to what happens to the level ℓ_t. The value of the slope at time t is thus a weighted average of the past value (b_{t-1}) and the "last observed slope", i.e. $\ell_t - \ell_{t-1}$. As for the level, the reason why a weighted average is chosen, is to compromise between the past of the time series and the most recent "fluctuation“.

In summary, the time series is described as a moving level (trend), which is continuously adapting over time, driven by a similarly adapting, time-dependent slope. The choice of using weighted averages amounts at defining the "scheme" (not to use the more inferential term "model"), defining the components.

Finally, the third equation states the way a new observation, h steps ahead in the future, is forecast, based on the past of the time series: by simply projecting the last level along the last slope, h steps in the future. Notice that "conditioning" on the entire past of the time series is the same as "conditioning" on the last available values of the level and the slope (in fact, to forecast the time series, we only need to know ℓ_T and b_T). This again shows that the level and the slope at time T summarize the entire past, given the THW scheme.

Error-correction form of the THW recursion. The THW recursion can be cast into an alternative form, called *error-correction*, which is of interest since it introduces, although still in a non-inferential way, the idea of an *error* acting in the computation of the hidden components. First of all, notice that the expression $\ell_{t-1} + b_{t-1}$ in the first THW equation is just the forecast of $\hat{y}_{t|\leq t-1}$ (which coincides with the forecast of the new level ℓ_t), i.e. the forecast of the time series at time t, given its past up to $t-1$. By putting $e_t = y_t - \hat{y}_{t|\leq t-1}$, the equation for the level can thus be rewritten as:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(y_t - \hat{y}_{t|\leq t-1}) = \ell_{t-1} + b_{t-1} + \alpha e_t. \tag{5}$$

This expression simply states that the level at time t is dynamically computed as its forecast from $t-1$, corrected by e_t (which incorporates the last observation). By substituting $\ell_t - \ell_{t-1} = b_{t-1} + \alpha e_t$ into the equation for the slope and rearranging, we finally get:

$$b_t = b_{t-1} + \alpha\beta e_t. \tag{6}$$

These two equations are the error-correction form of the THW recursion. In both cases, the components at time t are computed as the guess from the most recent past, plus a correction term, which includes the information carried in, by the last observation.

Setting smoothing parameters and initial conditions. To put the THW recursion to work, we need to set the smoothing parameters α and β and the initial conditions ℓ_2 and b_2, so as that the computations start from $t = 3$.

As for the Exponential smoothing procedure, the smoothing parameters can be tuned by minimizing the squared error $E^2(\alpha, \beta)$ on the observed time series, i.e. by minimizing the following quantity, viewed as a function of α and β:

$$E^2(\alpha, \beta) = \sum_{t=1}^{T} (y_t - \hat{y}_{t|\leq t-1})^2. \tag{7}$$

Initial conditions can be set in the following ways (which are heuristic in nature):

1. $\ell_2 = y_2$; $b_2 = y_2 - y_1$.
2. $\ell_2 = y_2$; $b_2 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3 = (y_4 - y_1)/3$.
3. Generalizing the previous choice, $\ell_2 = y_2$; $b_2 = (\sum_{i=1}^{m} y_i)/(m-1)$ $(m \geq 4)$.

Alternatively, the selection of initial values, can be performed jointly with that of the smoothing parameters, by minimizing the squared forecasting error, considered as a function of all the four parameters to set, i.e. by minimizing:

$$E^2(\alpha, \beta, \ell_2, b_2) = \sum_{t=3}^{T} (y_t - \hat{y}_{t|\leq t-1})^2. \tag{8}$$

As in the Exponential smoothing case, in order to detect possible overfitting, it is reccomandable to minimize the above error on a part of the time series (typically, limiting the sum to some $T' < T$) and checking the effectiveness of the resulting equations on the remaining part.

Remark. To simplify the THW model, sometimes one expresses both α and β as a function of a single parameter $0 < \omega < 1$. By setting $\alpha = 1-\omega^2$ and $b = (1-\omega)/(1+\omega)$, the error-correction form of the THW recursion becomes:

$$\begin{aligned} \ell_t &= \ell_{t-1} + b_{t-1} + (1-\omega^2)e_t & (9) \\ b_t &= b_{t-1} + (1-\omega)^2 e_t. & (10) \end{aligned}$$

This simplified form of the THW recursion is known as *double exponential smoothing*.

2.2 Seasonal Holt-Winters (SHW) method

We now introduce a seasonal component, into the THW recursion system; the equations of the SHW scheme read:

$$\begin{aligned} \ell_t &= \alpha(y_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) & (11) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} & (12) \\ S_t &= \gamma(y_t - \ell_t) + (1-\gamma)S_{t-s} & (13) \\ \hat{y}_{T+h|\leq T} &= \ell_T + hb_T + S_{T+h-ns}. & (14) \end{aligned}$$

where S_t is the seasonal component of the time series, s is its period (e.g., $s = 12$ for a monthly time series) and $0 < \alpha, \beta, \gamma < 1$.

Here, the equation for the level is modified, by considering the *de-seasonalized* time series $y_t - S_{t-s}$ in place of just the observation y_t. This is reasonable, since the level must not comprise the seasonal component. Notice that the seasonal component enters the equation for ℓ_t with time index $t - s$, i.e. we de-seasonalize the time series, by using the most recent guess for S_t, which is not S_{t-1}, but the value of the seasonal component computed one period before, i.e. S_{t-s}.

The equation for the slope is the same as that for the THW scheme, since the slope is not related to the seasonal component, in any way.

Finally, the recursion for the seasonal component computes S_t as a weighted average of the *de-trended* time series $y_t - \ell_t$ and the last guess S_{t-s}. Notice that here the level enters with time index equal to t.

The forecasting h steps ahead of the last observed time T is obtained by projecting the level along the slope h times ($\ell_T + hb_T$) and then adding the proper seasonal effect (the expression S_{T+h-ns} means "take the most recent guess for the seasonal component at time $T + h$, i.e. go back in time n steps from $T + h$, to the most recent corresponding season". It is indeed possible to give an explicit expression for n, but here we are just interested in understanding the meaning of the forecast relation).

Setting smoothing parameters and initial conditions, for the SHW scheme. As before, to run the SHW recursion, smoothing parameters and initial conditions must be set. Parameters are obtained by minimizing the squared error, as for the THW method. As to the initial conditions, one can heuristically set

- $\ell_s = \frac{1}{s}\sum_{i=1}^{s} x_i$
- $b_s = 0$.
- $S_i = x_i - \ell_s$, $i = 1, \ldots, s$.

The idea, is to initialize the level at time s, by the average of the first s observations and to compute the first s seasonal components as the de-trended time series (since the slope is initialized to 0, the level and the trend coincide).

As for the THW model, one can also set jointly the smoothing parameters and the initial conditions, by minimizing the forecasting error, considered as a function of all of the unknown quantities.

2.3 Seasonal damped Holt-Winters method

One of the main drawbacks of the THW and SHW schemes is that they forecast future observations in a very rigid way, i.e. by projecting the level along the slope, and possibly correcting with the proper seasonal component. It is the "projection on a straight line" to be quite unsatisfactory, being unrealistic for the slope to remain the same over time. To make forecasting somehow more rsensible, without really modifying the structure of the Holt-Winters schemes, a *damping factor* is introduced into the equations, which are then modified as follows:

$$\ell_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(\ell_{t-1} + \varphi b_{t-1}) \tag{15}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)\varphi b_{t-1} \tag{16}$$

$$S_t = \gamma(y_t - \ell_t) + (1-\gamma)S_{t-s} \tag{17}$$

$$\hat{y}_{T+h|\leq T} = \ell_T + (\varphi + \varphi^2 + \ldots + \varphi^h)b_T + S_{T+h-ns}, \tag{18}$$

where $0 < \varphi < 1$ is the damping factor mentioned above.

This factor enters explicitly the first and the second equation. In the first, we see that it dampens the value of the slope b_t, so that the "forecast level" $\ell_{t-1} + \varphi b_{t-1}$ (which enters the weighted average) is, in absolute value, less than the non-damped guess, used in the THW and SHW schemes. It enters in a similar way in the second equation, so that the evolution of the slope is linked to its damped past and not to its "pure" past (notice, in fact, that there is no damping factor in the left side of the second equation).

The last relation states the forecast equation for the damped scheme. Since $\varphi < 1$, $\varphi_h = \varphi + \varphi^2 + \ldots + \varphi^h < h$; moreover, as h increases, φ_h flattens, so the forecast is no more just a linear "projection" into the future. Clearly, the damping effect increases, as φ approaches to 0. To show why the forecasting relation has the form reported above, let us consider the case $h = 2$ and suppose we want to forecast $\hat{y}_{T+2|\leq T}$. If we knew ℓ_{T+1} and b_{T+1}, this would read $\hat{y}_{T+2|T} = \ell_{T+1} + \varphi b_{T+1} + S_{T+2-s}$ (consistently with the content of the recursion for ℓ_t, in the damped SHW scheme). Using the recursion for b_t in the last expression, we have:

$$\hat{y}_{T+2|\leq T} = \ell_{T+1} + \varphi(\beta(\ell_{T+1} - \ell_T) + (1-\beta)\varphi b_T) + S_{T+2-s}. \tag{19}$$

However, ℓ_{T+1} is unknown and so the best we can do is to substitute $\hat{\ell}_{T+1|\leq T} = \ell_T + \varphi b_T$ for it, getting:

$$\hat{y}_{T+2|\leq T} = \ell_T + \varphi b_T + \varphi^2 b_T + S_{T+2-s}. \tag{20}$$

For $h > 2$, the computations have the same structure.

The initialization and the choice of parameters value can be done analogously to the other schemes.

Remark. The *non-seasonal damped* Holt-Winters method is simply obtained by the previous system of equations, by putting $S_t = 0$ for all t and, consequently, removing the equation for the seasonal component, getting:

$$\begin{aligned} \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \varphi b_{t-1}) && (21)\\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\varphi b_{t-1} && (22)\\ \hat{y}_{T+h|\leq T} &= \ell_T + (\varphi + \varphi^2 + \ldots + \varphi^h) b_T. && (23) \end{aligned}$$

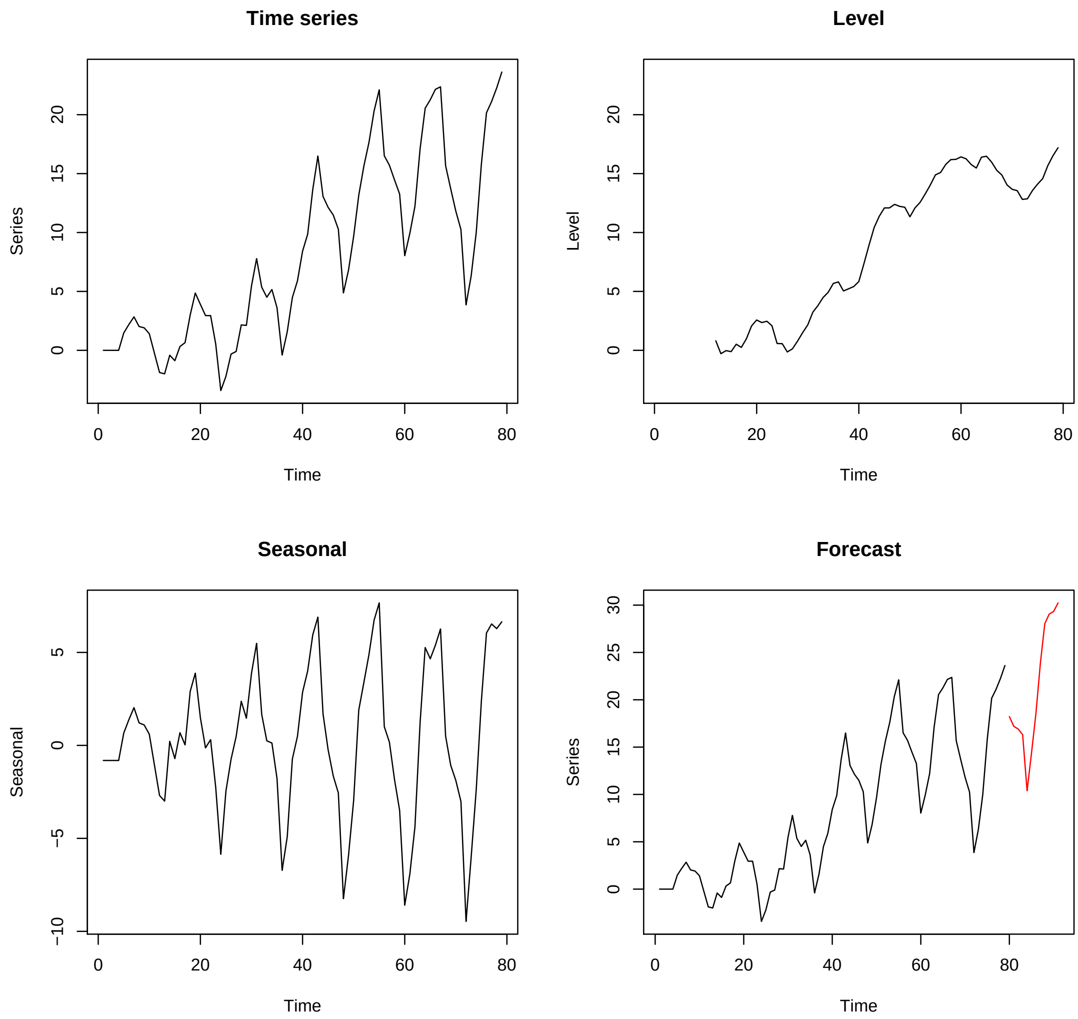

Figure 1: SHW decomposition and forecast (in red) of a monthly time series.