Financial Time Series and Their Characterictics

Mei-Yuan Chen

Department of Finance

National Chung Hsing University

Feb. 22, 2013

Contents

1 Introduction **1**

1.1 Asset Returns 1

1.2 Distributional Properties of Returns 4

1.2.1 Likelihood Function of Returns 8

1.2.2 Empirical Properties of Returns 8

1.3 Components of a Time Series 9

1.4 Time Series Smoothing 10

1.4.1 Smoothing via Moving Averages 10

1.4.2 Simple Exponential Smoothing 11

1.4.3 Holt's Linear Trend Algorithm 13

1.4.4 The Holt-Winter Algorithm 15

1.4.5 Other Exponential Smoothing Algorithms 17

1 Introduction

The future is neither completely knowable nor totally obscure; it is full of uncertainty. In our daily life we make forecasts from time to time, either implicitly or explicitly, and rely on these forecasts to make our decisions. We usually believe that the better the forecasts, the better will be the decisions.

There are numerous forecasting methods for different problems. We are primarily interested in the methods that can be justified scientifically. The behavior of a variable usually consists of a systematic component and an idiosyncratic component. The systematic part is characterized by a suitably constructed model from which forecasts can be obtained. A forecasting model is operational provided that it utilizes only the past information to generate forecasts. Owing to the presence of the idiosyncratic component, the resulting forecasts are not perfect in general. A variable without the systematic component can *not* be forecasted in a coherent manner. While we are learning different forecasting methods, we should keep in mind that these methods, more or less, have their own limitations.

1.1 Asset Returns

Instead of prices, asset returns are the objects of interest in financial studies. Two main reasons are raised by Campbell, Lo, and MacKinlay (1997). First, for general investors, return of an asset is a complete and scale-free summary of the investment opportunity. Second, asset prices are commonly observed empirically to be nonstationary which makes the statistical analysis difficult. There are several definitions of asset returns.

Let P_t be the price of an asset at time t. For the time being, no dividend being pad for the asset is assumed.

1. One-period Simple Returns: holding the asset for one period from date $t-1$to

t

(a) Simple Gross Return:

$$1 + R_{t,1} = \frac{P_t}{P_{t-1}}.$$

(b) Simple Net Returns (Simple Return):

$$R_{t,1} = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_t}.$$

2. Multiperiod Simple Return: holding the asset for one period from date $t-k$to t, the k-period simple gross return (or called a compound return) is

$$\begin{aligned} 1 + R_{t,k} = \frac{P_t}{P_{t-k}} &= \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} \\ &= (1 + R_{t,1})(1 + R_{t-1,1}) \cdots (1 + R_{t-k,1}) \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j,1}), \end{aligned}$$

and the k-period simple net return is

$$R_{t,k} = \frac{P_t - P_{t-k}}{P_{t-k}}.$$

3. Continuous Compounding Return: The effect of compounding can be illustrated with Table 1.1 in Tsay (2002):

Type	Number of payments	Interest rate per period	Net value
Annual	1	0.1	\$1.10000
Semiannual	2	0.05	\$1.10250
Quarterly	4	0.025	\$1.10381
Monthly	12	0.0083	\$1.10471
Weekly	52	0.1/52	\$1.10506
Daily	365	0.1/365	\$1.10516
Continuous	∞		\$1.10517

Above table summarizes the net values of a bank deposit \$1.00 with 10 % interest rate per annum for different times of interest payment in an year. For example, if the bank pays interest semi-annually, the 6-month interest rate is $0.1/2 = 0.05$ and the net value is $\$1.0(1+0.05)^2 = \$1.1025$ after the first year. In general, if the bank pays interest m times a year, then the interest rate of each payment is $0.1/m$ and the net value of the deposit becomes $\$1(1+0.1/m)^m$ one year later. As $m \to \infty$, $(1+0.1/m)^m = \exp(0.1)$ which is referred to as the *continuous compounding*.

In general, the net asset value A of continuous compounding is

$$A = C\exp(r \times n),$$

where r is the interest rate per annum, C is the initial capital, and n is the number of years. Then

$$C = A\exp(-r \times n)$$

is referred to as the *present value* of an asset that is worth A dollars n years from now.

The natural logarithm of the simple gross return of an asset is called the continuously compounded return or *log return*:

$$\begin{aligned} r_{t,1} &= \ln(1+R_{t,1}) = \ln\frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1} \\ &= p_t - p_{t-1}. \end{aligned}$$

As to the multiperiod returns, we have

$$\begin{aligned} r_{t,k} &= \ln(1+R_{t,k}) = \ln[(1+R_{t,1})(1+R_{t-1,1})\cdots(1+R_{t-k,1})] \\ &= \ln(1+R_{t,1}) + \ln(1+R_{t-1,1}) + \cdots + \ln(1+R_{t-k,1}) \\ &= r_{t,1} + r_{t-1,1} + \cdots + r_{t-k,1}. \end{aligned}$$

4. Portfolio Return: The simple net return of a portfolio consisting of N assets is a weighted average of the simple net returns of the assets involved, $R_{p,t} = \sum_{i=1}^{N} w_i R_{it}$. The weights are usually determined by the percentages of trading values (value weighted portfolio) and/or trading volumes (volume weighted portfolio) of the assets among total assets.

5. Dividend Payment: Suppose an asset pays dividend (D_t), periodically. The simple net return and continuously compound return at time t are

$$R_t = \frac{P_t + D_t}{P_{t-1}}, \qquad r_t = \ln(P_t + D_t) - \ln(P_{t-1}).$$

6. Excess Return: The excess return of an asset at time t is defined as the difference between it return and the return on some reference asset.

1.2 Distributional Properties of Returns

Consider a collection of N assets held for T periods. For each asset i, let r_{it} be the log return at time t. The log returns under study are $\{r_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$. The most general model for the log returns is their joint distribution function:

$$F_r(r_{11}, \ldots, r_{N1}; r_{12}, \ldots, r_{N2}; \ldots; r_{1T}, \ldots, r_{NT} | \boldsymbol{Y}, \theta),$$

where $\boldsymbol{Y}$ is a state vector consisting of variables that summarize the environment in which asset returns are determined and θ is a vector of parameters that uniquely determine the distributional function $F_r(\cdot)$. The probability distribution $F_r(\cdot)$ governs the stochastic behavior of the return r_{it} and $\boldsymbol{Y}$. In many financial studies, the state vector $\boldsymbol{Y}$ is treated as given and the main concern is the conditional distribution of $\{r_{it}\}$ given $\boldsymbol{Y}$.

Some financial theories such as CAPM focus on the joint distribution of N returns at a single time index t, i.e., $\{r_{1t}, r_{2t}, \ldots, r_{Nt}\}$. Other theories emphasize the dynamic structure of individual asset returns, i.e., $\{r_{i1}, \ldots, r_{iT}\}$. In the univariate time series analysis, our main concern is the joint distribution of $\{r_{it}\}_{t=1}^{T}$ for asset i. The joint distribution of $\{r_{it}\}_{t=1}^{T}$ can be partitioned as

$$\begin{aligned}
&F(r_{i1}, \ldots, r_{iT}; \theta) \\
&= F(r_{i1}; \theta) F(r_{i2}|r_{i1}; \theta) F(r_{i3}|r_{i1}, r_{i2}; \theta) \cdots F(r_{iT}|r_{i1}, \cdots, r_{i,T-1}; \theta) \\
&= F(r_{i1}; \theta) \prod_{t=2}^{T} F(r_{it}|r_{i,t-1}, \ldots, r_{i1}; \theta).
\end{aligned}$$

This partition highlights the temporal dependence of the log return r_{it}. The main issue then is the specification of the conditional distribution $F(r_{it}|r_{i,t-1}, \ldots, r_{i1})$ – in particular, how the conditional distribution evolves over time. The partition can also be represented in density functions:

$$f(r_{i1}, \ldots, r_{iT}; \theta) = f(r_{i1}; \theta) \prod_{t=2}^{T} f(r_{it}|r_{i,t-1}, \ldots, r_{i1}; \theta).$$

Several statistical distributions have been proposed in the literature for the marginal distributions of asset returns, including normal, lognormal, stable, and scale-mixture of normal distributions.

1. Normal Distribution: $\{R_{it}|t = 1, \ldots, T\}$ have been assumed as be independent and identically distributed as normal with fixed mean and variance. Drawbacks of this assumption are as follows. First, R_{it} has lower bound -1 however there is no bound for realizations of a normal distribution. Second, the multiperiod simple return $R_{it}[k]$ will not be normally distributed even R_{it} is normally distributed. Third, the normality assumption is not supported by most empirical empirical asset returns.

2. Lognormal Distribution: The log returns r_t of an asset is commonly assumed to be i.i.d. normally distributed with mean μ and variance σ^2. The simple

returns are then i.i.d. lognormal distributed with

$$E(R_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad \text{var}(R_t) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1].$$

Alternatively, let m_1 and m_2 be the mean and variance of the simple return R_t, which is lognormally distributed. The the mean and variance of the corresponding log return r_t are

$$E(r_t) = \ln\left[\frac{m_1 + 1}{\sqrt{1 + \frac{m_2}{(1+m_1)^2}}}\right], \quad \text{var}(r_t) = \ln\left[1 + \frac{m_2}{(1 + m_1)^2}\right].$$

As the log return r_t are i.i.d. normal, the multiperiod of log return $r_t[k]$ is also normally distributed. In addition, there is no lower bound for r_t, and the lower bound for R_t is satisfied using $1 + R_t = \exp(r_t)$. However, the lognormal assumption is not supported by the exhibition of a positive excess kurtosis in most asset returns.

3. Stable Distribution: r_t is stable iff its characteristic function h can be expressed as $h = e^g$, where g has one of the following forms: For $0 < \alpha < 1$ or $1 < \alpha \leq 2$,

$$g(u) = iu\delta - d|u|^\alpha \left(1 + i\beta\frac{u}{|u|}\tan(\frac{\pi}{2}\alpha)\right), \tag{1}$$

and for $\alpha = 1$,

$$g(u) = iu\delta - d|u| \left(1 + i\beta\frac{u}{|u|}\frac{2}{\pi}\ln|u|\right), \tag{2}$$

where $\delta \in R$, $d \geq 0$, $|\beta| \leq 1$, and take $u/|u| = 0$ when $u = 0$. Usually, equations (1) and (2) are called the characteristic function of the family of stable Paretian distribution. The parameter δ is a location parameter, d a scale parameter, β is a measure of of skewness, and α is the characteristic exponent. The characteristic exponent, α, determines the total probability in the extreme tails. The smaller the value of α, the thicker the tails of the distribution (Famma, 1963). The general form of the symmetric stable characteristic function located at zero, i.e., $\delta = 0, \beta = 0$, is

$$h(u) = \exp[-d|u|^{\alpha}], \qquad d \geq 0, \quad 0 < \alpha \leq 2.$$

When $\alpha = 2$, r_t is normal $(0, 2d)$; when $\alpha = 1$, r_t has the Cauchy density with parameter d.

If r_t is stable (not necessary symmetric) and $0 < \alpha \leq 1$, then h is not differentiable at $u = 0$, so $E(|r_t|) = \infty$. In the symmetric case, $E(r_t)$ does not exist. If $1 < \alpha < 2$, h can be differentiated once but not twice at $u = 0$, so that $E(r_t^2) = \infty$. This is to be expected, for if r_t has finite mean and variance, the fact that r_t can be obtained as a limit of a sequence of normalized sums implies that r_t must be normal. It can be shown that if r_t is stable, r_t has a finite rth moment for all $r \in (0, \alpha)$.

It is known that a normal random variable is a stable random variable with $\alpha = 2$, while a Cauchy is a stable random variable with $\alpha = 1$. Ash (1972, pp 345–346) pointed out that the normalized sum of i.i.d. Cauchy (special stable distribution with $\alpha = 1$) random variables has a limit which is also a Cauchy distribution. Moreover, the normalized sum of stable random variables has the same stable distribution as its limit. That means, the normalized sums of stable random variables will not follow the central limit theorem so that the functional central limit theorem breaks down. Due to its heavy dependence on the results of functional central limit theorem, the conventional large sample tests will be problematic in models with stable distributed errors.

4. Scale Mixture of Normal Distributions: An example of finite mixture of nor-

mal distributions is

$$r_t \sim (1-\alpha)N(\mu, \sigma_1^2) + \alpha N(\mu, \sigma_2^2),$$

where $0 \leq \alpha \leq 1$, σ_1^2 is small and σ_2^2

1.2.1 Likelihood Function of Returns

Suppose the conditional distribution $f(r_t | r_{t-1}, \ldots, r_1; \theta)$ (the subscript i is omitted) is normal with mean μ_t and variance σ_t^2, then $\boldsymbol{\theta}$ consists of the parameters in μ_t and σ_t^2 and the likelihood function of the data is

$$f(r_1, \ldots, r_T; \boldsymbol{\theta}) = f(r_1; \boldsymbol{\theta}) \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left[\frac{-(r_t - \mu_t)^2}{2\sigma_t^2}\right],$$

and the log likelihood function is

$$\ln f(r_1, \ldots, r_T; \boldsymbol{\theta}) = \ln f(r_1; \boldsymbol{\theta}) - \frac{1}{2}\sum_{t=2}^{T}\left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{(r_t - \mu_t)^2}{\sigma_t^2}\right].$$

1.2.2 Empirical Properties of Returns

1. Daily returns of the market indexes and individual stocks tend to have higher excess kurtoses than monthly returns. For monthly series, the returns of market indexes have higher excess kurtoses than individual stocks.

2. The mean of a daily return series is close to zero, whereas that of a monthly return series is slightly higher.

3. Monthly returns have higher standard deviations than daily returns.

4. Among the daily returns, market indexes have smaller standard deviations than individual stocks.

5. The skewness is not a serious problem for both daily and monthly returns.

6. The descriptive statistics show that the difference between simple and log returns is not substantial.

7. The empirical density function has a higher peak around it mean, but fatter tails than that of the corresponding normal distribution. In other words, the empirical density function is taller, skinner, but with a wider support than the corresponding normal density.

1.3 Components of a Time Series

As mentioned previously, a time series may be divided into a systematic component (a deterministic part and a stochastic part) and an idiosyncratic component. The deterministic part of the systematic component could be a deterministic function of time trend (including business cycle and time trend) and seasonality. The stochastic component consists of *autoregressive* and *moving average* elements. Furthermore, time-varying variance may also be an element of the stochastic component. To summary, a time series, y_t, can be represented as

$$\begin{aligned} y_t &= \{\text{systematic component}\} + \{\text{idiosyncratic component}\} \\ &= \{[\text{deterministic part}] + \text{stochastic part}\} + \{\text{idiosyncratic component}\} \\ &= \{[\text{business cycle} + \text{time trend}] + \text{seasonality}\} + \{\text{idiosyncratic component}\} \\ &= \{[f(t)] + g(s_t) + ARMA(p,q)\} + \sqrt{h_t}e_t. \end{aligned}$$

The aim of conventional time series analysis is to explore the functional forms of $f(t)$, $g(s_t)$, h_t and the orders of p and q. Tools of discovering $f(t)$, $g(s_t)$, h_t and p and q include the regression analysis, smoothing techniques, and the method of Box-Jenkins. For examples, $f(t)$ can be formulated as a linear $(\alpha_0 + \alpha_1 t)$ or quadratic $(\alpha_0 + \alpha_1 t + \alpha_2 t^2)$ function and $g(s_t)$ be modeled as $\gamma_1 s_{t1} + \gamma_2 s_{t2} + \gamma_3 s_{t_3}$

for quarterly time series, where $s_{ti}, i = 1, 2, 3$ are the seasonal dummy variables. As to the Box-Jenkins method, we will have complete discussion lately.

1.4 Time Series Smoothing

Smoothing techniques are ways of discovering time trend pattern of a time series. In other words, Smoothing techniques remove the jagged path of a time series.

1.4.1 Smoothing via Moving Averages

A time series usually exhibits a rather jagged time path so that its underlying regularities may be difficult to identify. To get a clearer picture of a time series, it is important to *smooth* its time path. A simple way of smoothing is to compute *moving averages* of the original series. Let y_t, $t = 1, \ldots, T$, be time series observations. The simple moving average with m periods is

$$y_t^* = \frac{y_{t-m+1} + \cdots + y_{t-1} + y_t}{m}, \qquad t = m, \ldots, T.$$

In technical analysis, y_t^* is usually taken as the 1-step ahead forecasting of y_{t+1} at time t. The 1-step ahead forecast error is defined as $e_t = y_{t+1} - y_t^*$. Observe that

$$\begin{aligned} y_t^* &= \frac{y_{t-m+1} + \cdots + y_{t-1} + y_t}{m} \\ y_{t-1}^* &= \frac{y_{t-m} + \cdots + y_{t-2} + y_{t-1}}{m} \end{aligned}$$

we have

$$\begin{aligned} y_t^* - y_{t-1}^* &= \frac{y_t - y_{t-m}}{m} \\ y_t^* &= y_{t-1}^* + \frac{y_t - y_{t-m}}{m}. \end{aligned}$$

This updating scheme makes the forecasting process much easier.

Notes:

1. Moving average is an easy and efficient way to understand and forecast the time path.

2. The drawback of using moving average is its inability to capture the peaks and troughs of the time series.

3. Under-prediction is obtained for data moving up persistently and over-prediction is for data moving down persistently.

4. The moving average is fail to deal with nonstationary time series.

5. Seasonality is eliminated by the moving average method.

6. Equal weight are given to all the data.

1.4.2 Simple Exponential Smoothing

A different approach of smoothing a time series is the so-called *exponential smoothing*. There are several exponential smoothing algorithms, each is constructed according to intuition, past experience, and certain characteristics of the time series under study. It is worth noting that this approach does not require fitting of a particular model.

We first discuss *simple exponential smoothing* which assigns a weight to the current observation y_t and exponentially decaying weights to previous observations as:

$$\begin{aligned} y_t^* &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \cdots \\ &= \alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{t-j}, \end{aligned}$$

where $0 < \alpha < 1$ is a *smoothing constant* to be determined by practitioners. As

$$\alpha \sum_{j=0}^{\infty} (1-\alpha)^j = 1,$$

y_t^* is a weighted average (linear combination) of current and past y_t and can be interpreted as an "estimate" of the current *level* of y_t. It is also easy to verify that y_t^* can be computed via the following simple algorithm:

$$y_t^* = \alpha y_t + (1-\alpha)y_{t-1}^*,$$

so that y_t^* is a linear combination of y_t and previously smoothed y_{t-1}^*. This algorithm typically starts with $y_1^* = y_1$. We do not discuss other choices of initial value because their effect on forecasts eventually vanish when t becomes large.

A drawback of simple exponential smoothing is that it yields constant forecasts for all future values. To see this, the forecast of y_{t+2} at $t+1$ is

$$y_{t+2|t+1} = y_{t+1}^* = \alpha y_{t+1} + (1-\alpha)y_t^*.$$

To make a 2-step ahead forecast, we may replace y_{t+1} by its forecast y_t^* and obtain

$$\begin{aligned} y_{t+2|t} &= y_{t+2|t+1|t} \\ &= [\alpha y_{t+1} + (1-\alpha)y_t^*]|t \\ &= \alpha y_{t+1|t} + (1-\alpha)y_t^* \\ &= \alpha y_t^* + (1-\alpha)y_t^* = y_t^*. \end{aligned}$$

Following the same line we find that the h-step ahead forecasts are $y_{t+h|t} = y_t$, $h = 1, 2, \ldots$. **Homework!** The *error-correction* form of the simple exponential smoothing algorithm is:

$$y_t^* = \alpha(y_{t-1}^* + e_t) + (1-\alpha)y_{t-1}^* = y_{t-1}^* + \alpha e_t.$$

This expression shows that positive (negative) forecast errors result in upward (downward) adjustments.

Another difficult problem associated with the simple exponential smoothing algorithm is the choice of smoothing constant α. An analyst may choose a smoothing constant subjectively based on his/her experience with similar time series. When the behavior of a time series is rather erratic so that an observation may contain a large irregular component, one would tend to adopt a smaller smoothing constant which gives less weight to the most recent observation but more weight to the previously smoothed estimate. For a smoother time series, a larger smoothing constant is then needed to give more weight to the most recent observation. This method relies on visual inspection of the time series; the exact weight to be assigned is determined quite arbitrarily. An objective way to determine a smoothing constant is the method of grid search. By selecting a grid of values for smoothing constant, we can compute sequences of smoothed series $y_t^*(\alpha)$ and their one-step forecast errors $e_t(\alpha)$. The "optimal" smoothing constant is the α for which the sum of squared one-step forecast errors, $\sum_{t=3}^{T} e_t(\alpha)^2$, is the smallest. Clearly, the effectiveness of this method depends on the choice of the grid.

Eviews demonstration

1.4.3 Holt's Linear Trend Algorithm

Writing $y_t = y_{t-1} + (y_t - y_{t-1})$, a better estimate of y_t may then be obtained by combining estimates of the level and change in level (local trend) of the underlying series. This motivates *Holt's linear trend algorithm*:

$$
\begin{aligned}
y_t^* &= \alpha y_t + (1-\alpha)(y_{t-1}^* + \tau_{t-1}), \\
\tau_t &= \beta(y_t - y_{t-1}^*) + (1-\beta)\tau_{t-1},
\end{aligned}
$$

where both α and β are smoothing constants between zero and one. This algorithm typically starts with $y_2^* = y_2$ and $\tau_2 = y_2 - y_1$. The algorithm can be expressed explicitly as

$$
\begin{aligned}
y_2^* &= y_2, \\
\tau_2 &= y_2 - y_1,
\end{aligned}
$$

$$\begin{aligned}
y_3^* &= \alpha y_3 + (1-\alpha)(y_2^* + \tau_2), \\
\tau_3 &= \beta(y_3 - y_2^*) + (1-\beta)\tau_2, \\
\vdots &= \vdots \\
y_T^* &= \alpha y_T + (1-\alpha)(y_{T-1}^* + \tau_{T-1}), \\
\tau_T &= \beta(y_T - y_{T-1}^*) + (1-\beta)\tau_{T-1}.
\end{aligned}$$

The forecast of y_{t+1}, $y_{t+1|t}$, is based on the current estimates of level and change in level, i.e., $y_{t+1|t} = y_t^* + \tau_t$. Hence,

$$y_{t+2|t+1} = y_{t+1}^* + \tau_{t+1} = \alpha y_{t+1} + (1-\alpha)(y_t^* + \tau_t) + \tau_{t+1}.$$

The 2-step ahead forecast is derived as:

$$\begin{aligned}
y_{t+2|t} &= y_{t+2|t+1|t} \\
&= \{[y_{t+1}^* + \tau_{t+1}]|t\} \\
&= \{\alpha y_{t+1} + (1-\alpha)(y_t^* + \tau_t) + \tau_{t+1}|t\} \\
&= \alpha y_{t+1|t} + (1-\alpha)(y_{t|t}^* + \tau_{t|t}) + \tau_{t+1|t} \\
&= \alpha y_{t+1|t} + (1-\alpha)(y_t^* + \tau_t) + \tau_t \\
&= \alpha(y_t^* + \tau_t) + (1-\alpha)(y_t^* + \tau_t) + \tau_t \\
&= y_t^* + 2\tau_t.
\end{aligned}$$

Similarly, the h-step ahead forecasts can be written as $y_{t+h|t} = y_t^* + h\tau_t$, $h = 1, 2, \ldots$. **Homework!!!**

In contrast with simple exponential smoothing, Holt's algorithm yields non-constant forecasts, but its projected future values grow (decline) by a fixed amount.

Let $e_t = y_t - y_{t-1}^* - \tau_{t-1}$ be the one-step forecast error. The error-correction form of Holt's algorithm becomes

$$\begin{aligned}
y_t^* &= \alpha(y_{t-1}^* + \tau_{t-1} + e_t) + (1-\alpha)(y_{t-1}^* + \tau_{t-1}) \;=\; y_{t-1}^* + \tau_{t-1} + \alpha e_t, \\
\tau_t &= \beta(y_t^* - y_{t-1}^*) + (1-\beta)\tau_{t-1} \;=\; \tau_{t-1} + \alpha\beta e_t.
\end{aligned}$$

Note that previous forecast errors affect both the estimates of level and local trend and that the adjustment of τ_t depends on α and β simultaneously.

To choose appropriate smoothing constants, we may still employ a grid search of pairs of values (α, β) to find the one minimizing the sum of squared one-step forecast errors. This method now must search for the best combination of two smoothing constants, and hence is computationally more demanding than for the simple exponential smoothing algorithm.

1.4.4 The Holt-Winter Algorithm

To allow for seasonality, we consider an extension of Holt's algorithm, which are known as the *Holt-Winters algorithm*. In particular, we consider both additive and multiplicative seasonality.

Let φ denote the seasonal factor and s its number of periods per year. Given additive seasonality, the Holt-Winters algorithm is

$$\begin{aligned} y_t^* &= \alpha(y_t - \varphi_{t-s}) + (1-\alpha)(y_{t-1}^* + \tau_{t-1}), \\ \tau_t &= \beta(y_t^* - y_{t-1}^*) + (1-\beta)\tau_{t-1}, \\ \varphi_t &= \gamma(y_t - y_t^*) + (1-\gamma)\varphi_{t-s}, \end{aligned}$$

where α, β, and γ are smoothing constants between zero and one. The first two equations are analogous to Holt's linear trend algorithm, except that the original series is first adjusted by subtracting the latest estimate of the seasonal factor φ_{t-s}. From the last equation we can see that a factor φ_t will not be used for updating until another s periods have elapsed. The initial values of this algorithm may be computed as: $\tau_s = 0$, $y_s^* = (y_1 + y_2 + \cdots + y_s)/s$, and $\varphi_i = y_i - y_s^*$, $i = 1, \ldots, s$.

It should be clear that, owing to additive seasonality, the forecasts of y_{t+h} are

$$y_{t+h|t} = \begin{cases} y_t^* + h\tau_t + \varphi_{t+h-s}, & h = 1, \ldots, s, \\ y_t^* + h\tau_t + \varphi_{t+h-2s}, & h = s+1, \ldots, 2s, \\ y_t^* + h\tau_t + \varphi_{t+h-3s}, & h = 2s+1, \ldots, 3s, \\ \vdots & \vdots \end{cases}$$

Note that each seasonal factor repeats every s periods. Let $e_t = y_t - y_{t-1}^* - \tau_{t-1} - \varphi_{t-s}$ be the one-step forecast error. The error-correction form of the Holt-Winter algorithm becomes:

$$\begin{aligned} y_t^* &= \alpha(y_{t-1}^* + \tau_{t-1} + e_t) + (1-\alpha)(y_{t-1}^* + \tau_{t-1}) \ = \ y_{t-1}^* + \tau_{t-1} + \alpha e_t, \\ \tau_t &= \beta(y_t^* - y_{t-1}^*) + (1-\beta)\tau_{t-1} \ = \ \tau_{t-1} + \alpha\beta e_t, \\ \varphi_t &= \gamma(y_t - y_t^*) + (1-\gamma)\varphi_{t-s}, \ = \ \varphi_{t-s} + \gamma(1-\alpha)e_t. \end{aligned}$$

Observe that the first two equations are the same as those of the Holt's algorithm, and the adjustment of φ_t also depends on α.

Similarly, given multiplicative seasonality, the Holt-Winters algorithm is:

$$\begin{aligned} y_t^* &= \alpha(y_t/\varphi_{t-s}) + (1-\alpha)(y_{t-1}^* + \tau_{t-1}), \\ \tau_t &= \beta(y_t^* - y_{t-1}^*) + (1-\beta)\tau_{t-1}, \\ \varphi_t &= \gamma(y_t/y_t^*) + (1-\gamma)\varphi_{t-s}. \end{aligned}$$

The initial values y_s^* and τ_s are the same as those for additive seasonality, and the initial values for seasonal factor are $\varphi_i = y_i/y_s^*$, $i = 1, \ldots, s$. The h-step ahead forecasts are:

$$y_{t+h|t} = \begin{cases} (y_t^* + h\tau_t)\varphi_{t+h-s}, & h = 1, \ldots, s, \\ (y_t^* + h\tau_t)\varphi_{t+h-2s}, & h = s+1, \ldots, 2s, \\ (y_t^* + h\tau_t)\varphi_{t+h-3s}, & h = 2s+1, \ldots, 3s, \\ \vdots & \vdots \end{cases}$$

Let $e_t = y_t - (y^*_{t-1} - \tau_{t-1})\varphi_{t-s}$. The error-correction form is:

$$\begin{aligned} y^*_t &= y^*_{t-1} + \tau_{t-1} + \alpha(e_t/\varphi_{t-s}), \\ \tau_t &= \tau_{t-1} + \alpha\beta(e_t/\varphi_{t-s}), \\ \varphi_t &= \varphi_{t-s} + \gamma(1-\alpha)(e_t/y^*_t). \end{aligned}$$

Although a grid search of appropriate smoothing constants is still plausible, it involves triples of values (α, β, γ) and is much more difficult to implement.

1.4.5 Other Exponential Smoothing Algorithms

A linear trend is not the only way to describe change in level of a time series. In this section we consider two other types of the trend component: damped trend and exponential trend. We describe only the algorithms for non-seasonal time series; their variants allowing for seasonality can be found in Newbold & Bos (1994).

In contrast with Holt's algorithm which predicts continuing growth, it may be more reasonable in some applications to predict that the growth of a time series eventually dies out. For example, given an estimated local trend τ_t, the predicted local trends may evolve as $c\tau_t$ at time $t+1$, $c^2\tau_t$ at $t+2$, and so on, where $0 < c \leq 1$ is the *damping factor*. The larger the damping factor, the slower the predicted trend diminishes. This leads to the *damped trend algorithm*:

$$\begin{aligned} y^*_t &= \alpha y_t + (1-\alpha)(y^*_{t-1} + c\tau_{t-1}), \\ \tau_t &= \beta(y^*_t - y^*_{t-1}) + (1-\beta)c\tau_{t-1}, \end{aligned}$$

and the h-step forecasts are:

$$y_{t+h|t} = y^*_t + \left(\sum_{j=1}^{h} c^j\right)\tau_t, \quad h = 1, 2, \ldots$$

Let $e_t = y_t - y^*_{t-1} - c\tau_{t-1}$. The error-correction form of this algorithm is:

$$\begin{aligned} y^*_t &= y^*_{t-1} + c\tau_{t-1} + \alpha e_t, \\ \tau_t &= c\tau_{t-1} + \alpha\beta e_t. \end{aligned}$$

Clearly, for $c = 1$, this algorithm simply reduces to Holt's algorithm.

In some other applications one may predict that future levels grow or decline exponentially over time. For notational simplicity, we now let τ_t denote growth rate, rather than local trend. The *exponential trend algorithm* is:

$$
\begin{aligned}
y_t^* &= \alpha y_t + (1-\alpha) y_{t-1}^* \tau_{t-1}, \\
\tau_t &= \beta (y_t^*/y_{t-1}^*) + (1-\beta)\tau_{t-1},
\end{aligned}
$$

and the h-step forecasts are:

$$
y_{t+h|t} = y_t^* \tau_t^h, \quad h = 1, 2, \ldots
$$

If there is a growth, i.e., $\tau_t > 1$, the predicted future values will increase exponentially with a constant growth rate. Let $e_t = y_t - y_{t-1}^* \tau_{t-1}$. The error-correction form of this algorithm is:

$$
\begin{aligned}
y_t^* &= y_{t-1}^* \tau_{t-1} + \alpha e_t, \\
\tau_t &= \tau_{t-1} + \alpha\beta (e_t / y_{t-1}^*).
\end{aligned}
$$