Chapter 2
Method for Constructing a Distribution-Free Index

Abstract Nonstationary financial time series often observed in the real world, include a time series with a slowly shifting mean value function, a time series with time-varying variations around the mean value, and a time series with both a moving mean value and changing waveforms around the mean value. First, we briefly review nonstationary time series modeling, such as trend estimation, time-varying variance modeling, seasonal adjustment modeling, and non-Gaussian distribution modeling, which is closely related to our method for constructing a distribution-free index. Since the distribution of prices of a financial market is often non-Gaussian, we propose to transform the price observations by the Box–Cox transformation. Then, a distribution-free index is defined by taking the inverse Box–Cox transformation of the optimal long-term trend, which is estimated by fitting a trend model with time-varying observation noises to the Box–Cox transformed observations. The new index becomes impartial, regardless of the price distributions.

Keywords Trend model · Nonstationary non-Gaussian time series · Time-varying variance · Distribution-free index · Box-Cox transformation

2.1 Nonstationary Time Series Modeling

2.1.1 Trend Estimation

Economic and financial time series often exhibit a slowly increasing or decreasing shift of the mean values over certain periods. We refer to a relatively long-term shift of the mean value as a trend. A trend may sometimes form a pattern due to an event specific to the attribute of the time series. These trends are important because economic and financial events influence all of our lives to a certain degree. Therefore, modeling a trend appropriately is important.

In order to capture the trend of a financial time series, for example, by calculating moving averages as in Fig. 2.1, which shows the 3-month moving averages of the Dow Jones Industrial Average, we practically attempt to draw a line or a curve as the trend. However, the delay of the point in time at which a trend changes cannot

Y. Tanokura and G. Kitagawa, *Indexation and Causation of Financial Markets*,
SpringerBriefs in Statistics, DOI 10.1007/978-4-431-55276-5_2

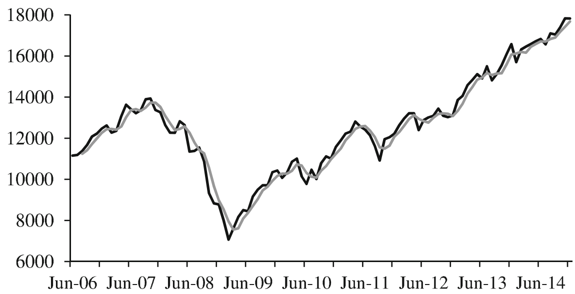


Fig. 2.1 Dow Jones Industrial Average (*black line*) and its 3-month moving averages (*gray line*). *Source* S&P Dow Jones Indices LLC

be ignored. This delay becomes longer as the period of taking moving averages becomes longer. In this book, we regard a trend as locally connected polynomials with stochastic fluctuations defined on short-term periods, which express gradual changes.

Next, consider a univariate observed time series, $y_n,\ n = 1, \ldots, N$, which is expressed as

$$y_n = t_n + w_n, \tag{2.1}$$

where t_n is referred to as the trend component, and w_n is a white noise following a Gaussian distribution with mean 0 and variance σ^2 (Kitagawa 2010). Then, y_n follows a Gaussian distribution with mean t_n and variance σ^2.

The trend component can be expressed in various forms (Kitagawa and Gersch 1984, 1996; Kitagawa 2010). Here, we define the trend component model as

$$\Delta^k t_n = v_n, \tag{2.2}$$

where k is the trend order, and v_n is a Gaussian white noise with mean 0 and variance τ^2. Δ is defined as the time difference operator satisfying

$$\Delta t_n = t_n - t_{n-1}. \tag{2.3}$$

We collectively refer to the pair of models (2.1) and (2.2) as the trend model. As the variance τ^2 of the noise v_n becomes smaller, the trend component model realizes smoother and more sensitive tendencies to the actual long-term fluctuations, as shown in Fig. 2.2.

When $k = 1$, (2.2) becomes a random walk model

$$t_n - t_{n-1} = v_n,$$

and the trend becomes locally constant. When $k = 2$, (2.2) becomes

$$t_n = 2t_{n-1} - t_{n-2} + v_n,$$

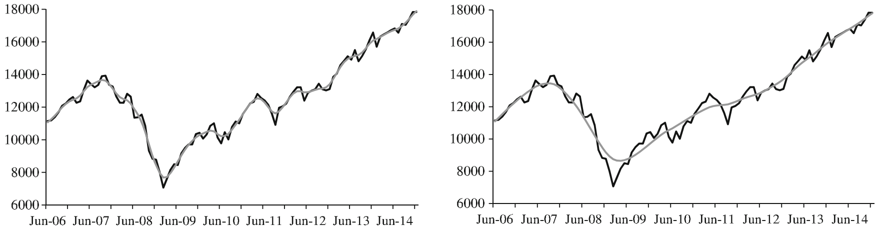


Fig. 2.2 Dow Jones Industrial Averages (*black*) and their estimated trends (*gray*). *Left-hand panel* $\tau^2 = 0.34$, *right-hand panel* $\tau^2 = 0.15$. *Source* S&P Dow Jones Indices LLC

and the trend is locally linear. Moreover, when $k = 3$, (2.2) is

$$t_n = 3t_{n-1} - 3t_{n-2} + t_{n-3} + v_n,$$

where the trend is locally quadratic. In the general case of k, by using the lag operator B defined by $B\ t_n = t_{n-1}$, the time difference operator of the kth order can be expressed as a binary expansion

$$\Delta^k = (1 - B)^k = \sum_{i=0}^{k} {}_kC_i(-B)^i. \tag{2.4}$$

Denoting the binomial coefficients $c_i = (-1)^{i+1}{}_kC_i$, the trend component model is written as

$$t_n = \sum_{i=1}^{k} c_i\ t_{n-i} + v_n, \tag{2.5}$$

which is formally an AR model. However, the trend component model is nonstationary because the roots of the characteristic equation lie on the unit circle.

The trend model given by (2.1) and (2.5) can be expressed by the following state-space model:

$$x_n = Fx_{n-1} + Gv_n \tag{2.6}$$
$$y_n = Hx_n + w_n, \tag{2.7}$$

where x_n is a k-dimensional state vector, F is a $k \times k$ matrix, and G and H are k-dimensional vectors defined by

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \\ \vdots \\ t_{n-k+1} \end{bmatrix}, \quad F = \begin{bmatrix} c_1 & c_2 & \cdots & c_k \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \tag{2.8}$$
$$H = [\, 1,\ 0,\ \ldots\ , 0\,].$$

The variances τ^2 and σ^2 of the noises are estimated by the maximum likelihood method, and the smoothed estimates of the state vector are calculated by the Kalman filter/fixed interval smoothing algorithm (Kitagawa 2010). The trend order k of the trend component model can be determined by the AIC (Akaike 1998; Konishi and Kitagawa 2008). Note that, in practice, the trend order k is usually selected as either 1 or 2.

When $k = 1$, the binomial coefficient in (2.5) is $c_1 = 1$, and the state-space model is obtained as

$$x_n = t_n, \quad F = G = H = 1. \tag{2.9}$$

When $k = 2$, $c_1 = 2$ and $c_2 = -1$ in (2.5), and the state-space model is obtained as

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, \ F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \ G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ H = \begin{bmatrix} 1, \ 0 \end{bmatrix}. \tag{2.10}$$

Moreover, the state-space model can also be obtained as

$$x_n = \begin{bmatrix} t_n \\ -t_{n-1} \end{bmatrix}, \ F = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}, \ G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ H = \begin{bmatrix} 1, \ 0 \end{bmatrix}. \tag{2.11}$$

We will use this representation (2.11) in the subsequent sections.

More details on the trend estimation can be found in Kitagawa and Gersch (1984, 1996) and Kitagawa (2010).

2.1.2 Time-Varying Variance Modeling

Nonstationary time series with time-varying fluctuations around the mean value can often be found in financial markets. In other words, the variance and the autocovariance function of such a time series change over time. In fact, as we sometimes hear the soar of the stock market volatility, such nonstationary phenomena always make us realize the existence of increased risk. Note that the estimation of a time-varying variance is equivalent to that of a stochastic volatility in financial time series analysis (Kitagawa 1987, 2010).

Estimating a time-varying variance directly is not easy. We estimate a time-varying variance using an approximated Gaussian distribution (Davis and Jones 1968) in the following manner. The advantage of this method is that the estimation of a time-varying variance can be realized by the simple estimation of the trend of a transformed time series.

Consider that a univariate time series, $y_n, \ n = 1, \ldots, N$, is the realization of a white noise that follows a Gaussian distribution with mean 0 and time-varying variance σ_n^2. We define the squared time series as follows:

$$s_n = y_n^2, \quad n = 1, \ldots, N. \tag{2.12}$$

Then, s_n follows a χ^2 (Chi-squared) distribution with one degree of freedom. Therefore, the probability density function of s_n is given by

$$f(s) = \frac{1}{\sqrt{2\pi\sigma^2}} s^{-\frac{1}{2}} \exp\left(-\frac{s}{2\sigma^2}\right). \tag{2.13}$$

Next, by the logarithmic transformation, we transform s_n to obtain

$$z_n = \log(s_n). \tag{2.14}$$

Then, since the inverse transformation of the logarithm is given by $s_n = e^{z_n}$, the probability density function of z_n is given by

$$g(z) = \left|\frac{de^z}{dz}\right| f(e^z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2}\left(z - \frac{e^z}{\sigma^2}\right)\right\}. \tag{2.15}$$

Since this $g(z)$ can be written as

$$g(z) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{1}{2}\{(z - \log\sigma^2) - \exp(z - \log\sigma^2)\}\right], \tag{2.16}$$

z_n can be expressed as

$$z_n = \log\sigma_n^2 + w_n. \tag{2.17}$$

The noise w_n in (2.17) follows a double exponential distribution, the probability density function of which is expressed as

$$h(w) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{1}{2}\{w - \exp(w)\}\right]. \tag{2.18}$$

The mean and the variance of this distribution are given by $-(\gamma + \log 2) = -(0.57722 + 0.69315) = -1.27036$ (γ: Euler constant) and $\pi^2/2$, respectively.

Therefore, by approximating the double exponential distribution as a Gaussian distribution with mean $-(\gamma + \log 2)$ and variance $\pi^2/2$, the estimation of the logarithm of the variance σ_n^2 of the original time series y_n can be reduced to that of the following trend model:

$$\Delta^k t_n = v_n \tag{2.19}$$

$$z_n = t_n + w_n, \tag{2.20}$$

where k is the trend order, and the system noise v_n follows a Gaussian distribution with mean 0 and variance τ^2.

By applying state-space modeling as described in the previous section, the trend component t_n is estimated by the Kalman filter/fixed interval smoothing algorithm (Kitagawa 2010). Since the smoothed estimates of $\log\sigma_n^2$ are obtained by

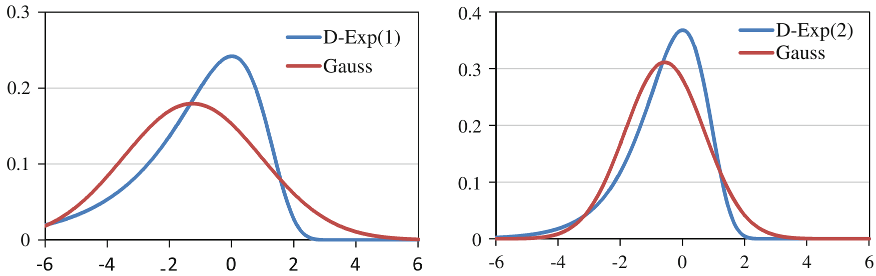


Fig. 2.3 Double exponential (D-Exp) distributions (*blue*) and their Gaussian approximations (*red*). *Left-hand panel* one degree of freedom, *Right-hand panel* two degrees of freedom

$$t_n + \gamma + \log 2, \quad n = 1, \ldots, N, \tag{2.21}$$

$\exp(t_n + \gamma + \log 2)$ is the estimate of the time-varying variance.

However, as shown in the left panel of Fig. 2.3, the double exponential distribution derived from the χ^2 distribution with one degree of freedom (blue line) is highly skewed and the variance is large, and so the approximation by the Gaussian distribution (red line) is not so good.

Therefore, in order to mitigate this problem in the actual estimation of the time-varying variance, we usually define the following time series:

$$s_m = \frac{1}{2}\left(y_{2m-1}^2 + y_{2m}^2\right), \quad m = 1, \ldots, N/2. \tag{2.22}$$

Then, s_m follows a χ^2 distribution with two degrees of freedom, i.e., an exponential distribution. Using a similar argument to that above, the density function of the logarithm of (half of) the χ^2 distribution with two degrees of freedom is given by

$$g(z) = \frac{1}{\sigma^2}\exp\left(z - \frac{e^z}{\sigma^2}\right) = \exp\left\{(z - \log\sigma^2) - \exp(z - \log\sigma^2)\right\}. \tag{2.23}$$

In this case, the noise w_m, replacing n with m in (2.17), follows a double exponential distribution in which the probability density function is expressed as

$$h(w) = \exp\{w - \exp(w)\}. \tag{2.24}$$

The mean and the variance are given by $-\gamma = -0.57722$ (γ: Euler constant) and $\pi^2/6$, respectively, as shown in the right panel of Fig. 2.3.

In this method of time-varying variance estimation, the number of the observations is halved. On the other hand, the noise distribution $h(w)$ in (2.24) is closer to a Gaussian distribution than that in the original (2.18), and the variance of the observation noise becomes a third of that of the original approximation. Therefore, if the variance does not change abruptly over time and the assumption that $\sigma_{2m-1}^2 = \sigma_{2m}^2$ is

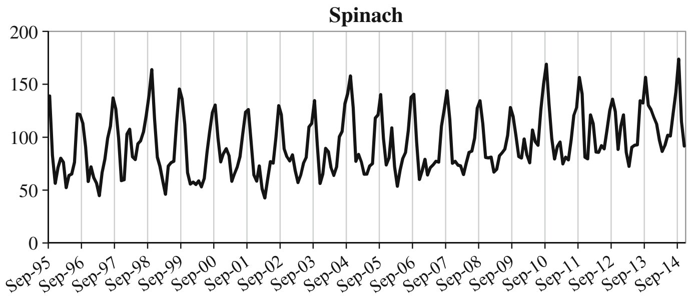


Fig. 2.4 Japanese consumer price index for spinach. *Source* Statistics Bureau of Japan

reasonable, the accuracy of the estimation of $\log \sigma_n^2$ can be expected to increase by approximately 50 %, i.e., the variance becomes $2 \times 1/3 = 2/3$. Several examples of the estimation of a time-varying variance are presented in Chap. 4.

Note that direct estimation can be performed by applying a non-Gaussian filter/smoothing algorithm without the assumption of $\sigma_{2m-1}^2 = \sigma_{2m}^2$ (Kitagawa 2010).

2.1.3 Seasonal Adjustment Modeling

Some economic time series that are closely related to financial markets tend to reflect seasonal factors, i.e., exhibit a similar pattern of fluctuations around the same season every year. Familiar examples of such time series are the prices of vegetables, wages, and unemployment rates. Figure 2.4 shows the Japanese consumer price index for spinach. A clear seasonal pattern with a yearly peak in September exists. These time series may influence not only financial markets but also economies.

This section briefly introduces a seasonal adjustment model proposed by Gersch and Kitagawa (1983), and Kitagawa and Gersch (1984, 1996).

When a component s_n of a time series cyclically fluctuates on a yearly basis, this component can be expressed as

$$s_n \approx s_{n-p}, \tag{2.25}$$

where p is the period length of the component. For example, in the case of monthly data, set $p = 12$. Then, s_n is referred to as a seasonal component or seasonality, which has a more or less regular fluctuation with a period of 1 year. Note that this seasonal component can be applied to other regular patterns, such as the weekly pattern in daily data ($p = 7$) or the daily pattern in hourly data ($p = 24$).

For an observed univariate time series y_n with a regular seasonal pattern of fluctuations, a seasonal adjustment model is expressed as

$$y_n = t_n + s_n + p_n + w_n. \tag{2.26}$$

In other words, the observation comprises the following four components.

A trend component t_n is estimated by the trend component model with the trend order k

$$\Delta^k t_n = v_{n1}, \quad v_{n1} \sim N(0, \tau_1^2), \tag{2.27}$$

which was presented in the previous section.

A seasonal component s_n, which slowly forms seasonal fluctuations, is expressed by the following seasonal component model with a period length p:

$$\sum_{i=0}^{p-1} s_{n-i} = v_{n2}, \quad v_{n2} \sim N\left(0, \tau_2^2\right). \tag{2.28}$$

The details of this model will be explained later.

A stationary component p_n is estimated by the following stationary AR component model of order m:

$$p_n = \sum_{i=1}^{m} a_i p_{n-i} + v_{n3}, \quad v_{n3} \sim N(0, \tau_3^2), \tag{2.29}$$

which expresses relatively shorter cyclical fluctuations than the gradual long-term trend component (2.27). Finally, the distribution of the observation noise w_n in (2.26) is given by

$$w_n \sim N(0, \sigma^2). \tag{2.30}$$

Economic time series with seasonality often accompany a trend expressing gradually shifting mean value functions, such as the consumer price index. For such series, the observation model is naturally considered as a form of decomposition: a trend component plus a seasonal component. In order to obtain a smoother trend, we introduce a seasonal adjustment model to which a stationary AR component expressing shorter cyclical fluctuations than a trend is added.

Let us now explain the seasonal component model (2.28). Using the lag operator B introduced in Sect. 2.1.1, we obtain

$$B^p s_n = s_{n-p}. \tag{2.31}$$

Therefore, (2.25) approximately satisfies

$$(1 - B^p)s_n \approx 0. \tag{2.32}$$

Similar to the trend component model presented in Sect. 2.1.1, a seasonal component with period p, can approximately be defined as

$$\left(1 - B^p\right) s_n = v_{n2}, \tag{2.33}$$

where v_{n2} is a white noise following a Gaussian distribution with mean 0 and unknown variance τ_2^2.

However, in practice, this seasonal component model (2.33) may not work well within the framework of the seasonal adjustment model (2.26) due to the existence of the factor $(1 - B)$, which is common to both the trend and seasonal component models. Since, as compared with the lag operator expression (2.4) of the trend model in Sect. 2.1.1, the following expansion

$$1 - B^p = (1 - B)(1 + B + \cdots + B^{p-1}), \tag{2.34}$$

is obtained for the seasonal component model (2.33).

Next, any arbitrary constant $e_n = c$ satisfies the difference equation

$$(1 - B)e_n = 0. \tag{2.35}$$

Therefore, if we define other components t_n' and s_n' as

$$\begin{aligned} t_n' &= t_n + e_n \\ s_n' &= s_n - e_n, \end{aligned}$$

then these components satisfy (2.27), (2.33), and

$$y_n = t_n' + s_n' + p_n + w_n. \tag{2.36}$$

Therefore, apart from the stationary AR component p_n, we have two methods by which to decompose the time series y_n into t_n and s_n with the same noises v_{n1}, v_{n2}, and w_n, and there is nothing to choose between them. Using the common factor in the component models of t_n and s_n within the seasonal adjustment modeling framework, the uniqueness of decomposition is lost.

From the expansion (2.34), as the sufficient condition for $1 - B^p = 0$ is $1 + B + \cdots + B^{p-1} = 0$, when $\sum_{i=0}^{p-1} B^i s_n \approx 0$ is satisfied,

$$s_n \approx s_{n-p}$$

is also satisfied. Therefore, in order to avoid the above problem of the nonuniqueness of the decomposition, define the following seasonal component model:

$$\sum_{i=0}^{p-1} B^i s_n = v_{n2}, \quad v_{n2} \sim N\big(0, \tau_2^2\big). \tag{2.37}$$

Therefore, (2.28) is obtained.

Equivalently, since the seasonal component model (2.37) can be written as

$$s_n = -\sum_{i=1}^{p-1} B^i s_n + v_{n2}, \tag{2.38}$$

this model can formally be regarded as a special case of an AR model.

As in the case of the trend model, the state-space model is obtained as

$$x_n = \begin{bmatrix} s_n \\ s_{n-1} \\ \vdots \\ s_{n-p+2} \end{bmatrix}, \quad F = \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$
$$H = [\ 1,\ 0,\ \ldots,\ 0\ ]. \tag{2.39}$$

The details and an extension to a higher seasonal order model can be found in Kitagawa (2010).

Considering the models mentioned above, we now return to the seasonal adjustment model (2.26). Each component model can be expressed in state-space model form. Therefore, in the same manner as for each component model, the state-space model used for the seasonal adjustment model (2.26) consisting of (2.27)–(2.29) is obtained in the following composite form:

$$x_n = \begin{bmatrix} x_{1n} \\ x_{2n} \\ x_{3n} \end{bmatrix}, \quad F = \begin{bmatrix} F_1 & & \\ & F_2 & \\ & & F_3 \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \end{bmatrix},$$
$$H = [\, H_1\ H_2\ H_3 \,], \quad Q = \begin{bmatrix} \tau_1^2 & 0 & 0 \\ 0 & \tau_2^2 & 0 \\ 0 & 0 & \tau_3^2 \end{bmatrix}. \tag{2.40}$$

Here, F_i, G_i, and H_i correspond to the matrices in the state-space representation for each component model. Similarly, Q is composed of the set of the variances of the system noises.

The above-mentioned seasonal adjustment model is freely available as web-based time series analysis software, Web DECOMP, at http://ssnt.ism.ac.jp/inets/inets_eng.html, which was developed by the Institute of Statistical Mathematics. In Chap. 4, DECOMP is used to extract a detrended cyclical component of a time series in order to detect causations between short-term fluctuations of financial and economic time series by power contribution analysis.

As an example of applying the seasonal adjustment model provided by DECOMP, Fig. 2.5 shows a decomposition of the Japanese consumer price index for spinach illustrated in Fig. 2.4, into the trend, seasonal, stationary AR, and noise components. The highly visible seasonality is detected.

2.1.4 Non-Gaussian Distribution Modeling

In the Gaussian state-space modeling presented in the previous sections, the gradual changes of fluctuation structures of nonstationary time series are well captured. However, in financial time series, time-varying fluctuation structures, occasionally

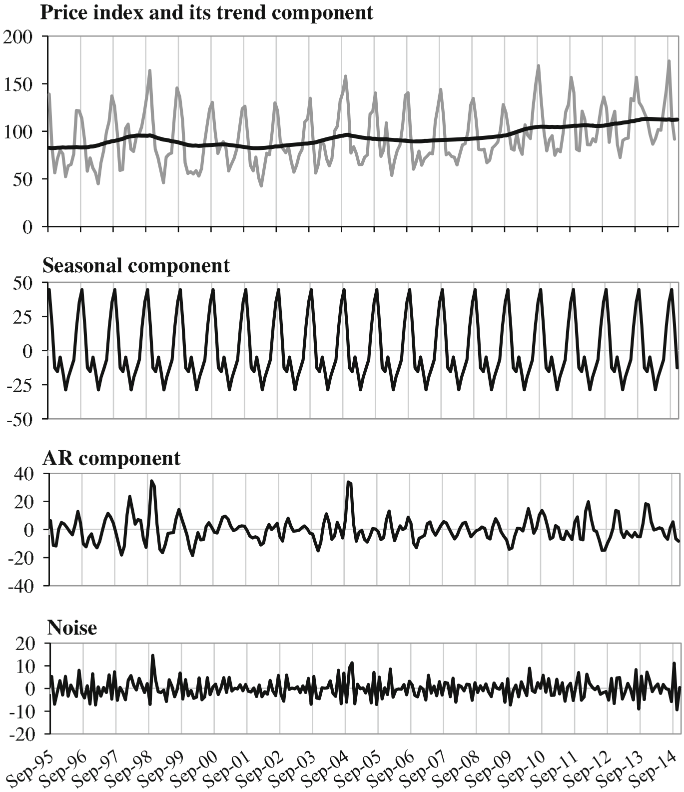


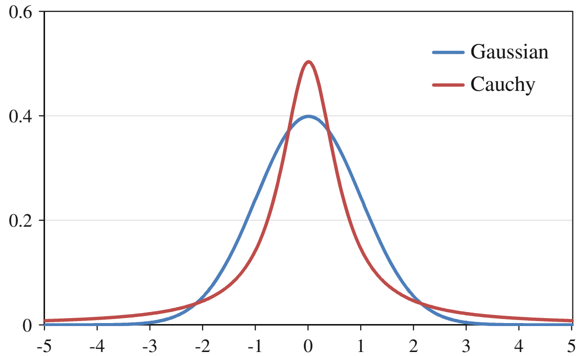
Fig. 2.5 Decomposition of the Japanese consumer price index for spinach (*gray line*) into the trend (*top*), seasonal (second from *top*), AR (second from *bottom*), and noise (*bottom*) components. *Source* Statistics Bureau of Japan

including both gradual and sudden changes, can often be observed. The recent outstanding example of a sudden change can be the occurrence of the Lehman Brothers' bankruptcy in 2008. Actually, the possibility of a sudden drastic change has recently expanded and recognizing indicators of such change are crucial.

As an extension of Gaussian state-space modeling, a non-Gaussian state-space model is naturally considered, by assuming a non-Gaussian distribution of the system noise, a non-Gaussian distribution of the observation noise, or non-Gaussian distributions of both. For example, when a heavy-tailed distribution, such as the Cauchy distribution (red line) shown in Fig. 2.6, is assumed for the system noise, a time series including gradual changes with high probabilities and sudden changes with low probabilities, can be modeled. Furthermore, note that the estimation of a time-varying variance in Sect. 2.1.2 can be directly performed using the noise distribution (2.18) without using the Gaussian approximation.

In order to extend the standard state-space model (1.14) and (1.15) mentioned in the introductory chapter, we recall again that

Fig. 2.6 Gaussian (*blue*) and Cauchy (*red*) distributions



$$x_n = F x_{n-1} + G v_n \tag{2.41}$$

$$y_n = H x_n + w_n, \tag{2.42}$$

where x_n is the state vector at time n. Here, the density functions $q(v)$ and $r(w)$ of the system noise v_n and the observation noise w_n are not necessarily Gaussian. Then, the state distribution generally becomes non-Gaussian.

In Gaussian state-space modeling, since the state distribution is Gaussian, the conditional means and the conditional variance covariance matrices are obtained recursively by the Kalman filter. However, in non-Gaussian state-space modeling, as the conditional distribution of the state cannot be specified only by the conditional mean and the conditional variance covariance matrix, computing the state distribution is necessary.

In order to address this problem, various algorithms, such as the extended Kalman filter (Anderson and Moore 2012), and the non-Gaussian filter/smoothing algorithm which numerically approximates non-Gaussian distributions by using a step function or a piecewise linear function (Kitagawa 1987), have been proposed. The development of various non-Gaussian approximating algorithms is reported in Kitagawa and Gersch (1996).

Denoting the set of observations by time t as $Y_t \equiv \{y_1, \ldots, y_t\}$, in general for the non-Gaussian state-space model (2.41) and (2.42), the conditional distribution of the state $p(x_n|Y_t)$ is obtained by the following recursive formula:

$$\begin{aligned} p(x_n|Y_{n-1}) &= \int p(x_n|x_{n-1}) p(x_{n-1}|Y_{n-1}) dx_{n-1} \\ p(x_n|Y_n) &= \frac{p(y_n|x_n) p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})}, \end{aligned} \tag{2.43}$$

where $p(y_n|Y_{n-1}) = \int p(y_n|x_n) p(x_n|Y_{n-1}) dx_n$.

Using this non-Gaussian filter/smoothing algorithm, it is possible to estimate a time-varying variance of a time series without approximating the double exponential distribution by a Gaussian distribution, i.e., by directly using the time-varying variance model

$$
\begin{aligned}
t_n &= t_{n-1} + v_n \\
x_n &= t_n + w_n,
\end{aligned}
\tag{2.44}
$$

where the trend order is one, and the noise distribution is expressed as $h(w_n) = (2\pi)^{1/2} \exp\{(w_n - \exp(w_n))/2\}$ in Sect. 2.1.2. Moreover, note that in this method, the system noise v_n is not restricted to be Gaussian. By using a heavy-tailed distribution for the system noise, sudden changes in variance can be detected (Kitagawa 1987, 2010).

This non-Gaussian filter/smoothing algorithm can also be applied to a nonlinear state-space model

$$
\begin{aligned}
x_n &= f(x_{n-1}, v_n) \\
y_n &= h(x_n, w_n).
\end{aligned}
\tag{2.45}
$$

In this nonlinear modeling framework, the time-varying variance model can be expressed as

$$
\begin{aligned}
s_n &= s_{n-1} + v_n \\
y_n &= e^{s_n} w_n.
\end{aligned}
\tag{2.46}
$$

Therefore, using this method, we can estimate the time-varying variance, even without defining the squared time series, as in Sect. 2.1.2 (Kitagawa 2010).

For the more general case, Kitagawa (1996) proposed a significantly practical simulation-based estimation method, i.e., the sequential Monte Carlo filter, for a nonlinear non-Gaussian state-space representation further extending to nonlinearities of the state and (or) the observation models. This filter approximates a distribution by several (for example, 10,000 or more) particles that can be regarded as independent realizations from the distribution (Gordon et al. 1993; Kitagawa 1996, 2010; Doucet et al. 2001).

Next, we briefly outline the sequential Monte Carlo filter. For the set of observations by time t, Y_t, we will evaluate the conditional distribution of the state $p(x_n|Y_t)$, which is referred to as a predictor when $n > t$, as a filter when $n = t$, and as a smoother when $n < t$.

The initial state x_0 is assumed to follow the density $p_0(x)$, and, for the above three cases, each distribution is expressed using m particles, as follows:

$$
\begin{aligned}
\{p_n^{(1)}, \ldots, p_n^{(m)}\} &\sim p(x_n|Y_{n-1}) \quad &&\text{for the predictor,} \\
\{f_n^{(1)}, \ldots, f_n^{(m)}\} &\sim p(x_n|Y_n) \quad &&\text{for the filter,} \\
\{s_{n|t}^{(1)}, \ldots, s_{n|t}^{(m)}\} &\sim p(x_n|Y_t) \quad &&\text{for the smoother.}
\end{aligned}
$$

When m particles $\{p_n^{(1)}, \ldots, p_n^{(m)}\}$ from the predictor $p(x_n|Y_{n-1})$ are given, the distribution is approximated by the empirical distributions determined by the m particles. In other words, the distribution is approximated by the probability mass function

$$\Pr\left(x_n = p_n^{(j)} | Y_{n-1}\right) = \frac{1}{m}, \quad j = 1, \ldots, m.$$

Then, a set of realizations expressing the one step ahead predictor $p(x_n|Y_{n-1})$ and the filter $p(x_n|Y_n)$ can be obtained recursively in the following manner:

1. Generate a random number $f_0^{(j)} \sim p_0(x)$ for $j = 1, \ldots, m$.
2. Repeat the following steps for $n = 1, \ldots, t$.

   a. Generate random numbers $v_n^{(j)} \sim q(v)$ for $j = 1, \ldots, m$, to obtain independent realizations of the system noise v_n in (2.41) following the distribution with density function $q(v)$.
   b. Compute $p_n^{(j)} = F f_{n-1}^{(j)} + G v_n^{(j)}$ for $j = 1, \ldots, m$.
   c. Compute $\alpha_n^{(j)} = r(y_n - H p_n^{(j)})$ for $j = 1, \ldots, m$, where $r(w)$ is the density function of the observation noise w_n in (2.42).
   d. Generate $f_n^{(j)}$ for $j = 1, \ldots, m$, by resampling $p_n^{(1)}, \ldots, p_n^{(m)}$ with weights proportional to $\alpha_n^{(1)}, \ldots, \alpha_n^{(m)}$.

For a parameter set θ of the state-space model such as the variances of the noises, the likelihood of the model is given by

$$L(\theta) = p(y_1, \cdots, y_t|\theta) = \prod_{n=1}^{t} p(y_n|Y_{n-1}),$$

where $p\,(y_1|Y_0) = p_0(y_1)$. For applying the sequential Monte Carlo filter, we use the approximation

$$\begin{aligned} p\,(y_n|Y_{n-1}) &= \int p(y_n|x_n)p(x_n|Y_{n-1})dx_n \\ &\cong \frac{1}{m}\sum_{j=1}^{m} p\left(y_n|p_n^{(j)}\right) = \frac{1}{m}\sum_{j=1}^{m} \alpha_n^{(j)}. \end{aligned}$$

The maximum likelihood estimate can be obtained by maximizing the log-likelihood:

$$\log L(\theta) = \sum_{n=1}^{t} \log p(y_n|Y_{n-1}) \cong \sum_{n=1}^{t} \log\left(\sum_{j=1}^{m} \alpha_n^{(j)}\right) - t \log m.$$

The sequential Monte Carlo filter is described in detail in Kitagawa (2010), Kitagawa and Gersch (1996), and Doucet et al. (2001).

The application of the sequential Monte Carlo filter is shown in Chap. 4.

2.2 Transformation of Non-Gaussian Distributed Prices of a Financial Market

Financial markets are discussed daily in the news throughout the world. The movement of some markets can influence business and economy, and such markets may replace each other over time. Since we usually observe financial market indices such as the S&P 500, one means by which to express the overall perspective of a market is to use the index as a proxy measure. Unlike an established market, in which the index is officially defined and announced, for a newly developed financial instrument forming its market with rapid growth, it is not easy to construct an appropriate index due to a lack of information, such as missing observations at certain times. Moreover, in order to fully reflect the price movements of a financial asset, the index should reflect the price distributions.

Although the assumption that the distributions of prices or returns of financial assets are Gaussian has been commonly used in theoretical finance, some studies in the 1960s failed to validate this assumption and found heavier tails than would be present in a Gaussian distribution (Mandelbrot 1963; Fama 1965). The tails consisting of extreme values of prices or returns that are caused by sharply soaring or plunging asset prices are more likely to occur than expected by a Gaussian distribution. In particular, distributions of stock returns have been discussed in many studies such as Paraez (1972), Madan and Seneta (1990), and Linden (2001). However, an exact identification of such distributions remains an open question.

The distribution of Credit Default Swap (CDS) spreads is often significantly heavy-tailed. For example, Fig. 2.7 shows histograms of the Japanese corporate CDS spreads referencing 327 companies. From back to front, the distributions are heavily skewed to the right. An analysis of this market is provided in Chap. 4.

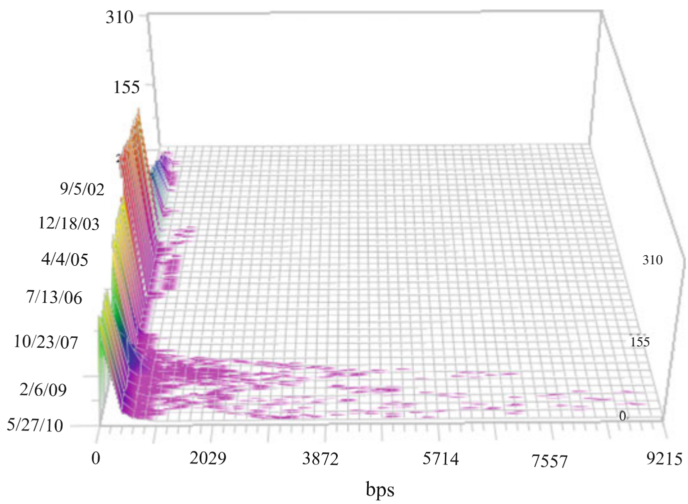


Fig. 2.7 Japanese corporate CDS spread histograms. *Source* Bloomberg LP

In order to facilitate the identification of such a distribution, Tanokura et al. (2012) proposed transforming the observations in the following manner:

Let $p_i(n),\ i = 1, \ldots, j(n)$, denote the prices of issues of a financial market with a non-Gaussian price distribution at time $n,\ n = 1, \ldots, N$. The number of observations $j(n)$ varies over time and can be zero at certain times, and $p_i(n)$ is positive. In order to transform a skewed non-Gaussian distribution of the prices to an approximately Gaussian distribution, we consider the Box–Cox transformation (Box and Cox 1964):

$$q_{i,\lambda}(n) = h(p_i(n)) = \begin{cases} \lambda^{-1}\{p_i(n)^{\lambda} - 1\} & \lambda \neq 0 \\ \log\ p_i(n) & \lambda = 0. \end{cases} \tag{2.47}$$

This transformation has been applied in various areas of finance and includes most major transformations, as well as no transformation, as follows: Ignoring a constant term, the Box–Cox transformation becomes the inverse transformation for $\lambda = -1$, the reciprocal square root transformation for $\lambda = -0.5$, the logarithm for $\lambda = 0$, the square root for $\lambda = 0.5$, and no transformation for $\lambda = 1$.

Now, for each λ, since there are $j(n)$ observations at time n, consider the following average time series of the Box–Cox transformed prices $q_{i,\lambda}(n)$:

$$y_\lambda(n) = \frac{1}{j(n)} \sum_{i=1}^{j(n)} q_{i,\lambda}(n), \quad n = 1, \ldots, N, \tag{2.48}$$

which is often observed to be nonstationary. Then, we fit the following trend model: to $y_\lambda(n)$:

$$\Delta^k t_\lambda(n) = v_\lambda(n), \qquad v_\lambda(n) \sim N(0, \tau_\lambda^2) \tag{2.49}$$

$$y_\lambda(n) = t_\lambda(n) + w_\lambda(n), \qquad w_\lambda(n) \sim D(0, \sigma_\lambda^2), \tag{2.50}$$

where k is the trend order, and $\Delta t_\lambda(n) = t_\lambda(n) - t_\lambda(n-1)$. Here, $D(0, \sigma_\lambda^2)$ denotes a general distribution with location parameter 0 and unknown scale parameter σ_λ.

This is an extension of the trend model with the Gaussian observation noises shown in Sect. 2.1.1 to the trend model with general observation noises. In other words, the case of a non-Gaussian observation noise distribution can also be considered.

In order to consider a rapidly growing or immature financial market with a significantly changing number of observations over time, it might be reasonable to assume that σ_λ^2 in (2.50) is inversely proportional to the number of observations. In other words, we replace (2.50) in the above trend model with the following expression:

$$y_\lambda(n) = t_\lambda(n) + w_\lambda(n), \qquad w_\lambda(n) \sim D(0, \sigma_\lambda^2 / j(n)). \tag{2.51}$$

Tanokura et al. (2012) considered the trend model with Cauchy observation noises in (2.51), which is generally useful for modeling the large deviation of noises that are often observed in financial markets. For convenience, we refer to this trend estimation model based on Cauchy observation noises as the Cauchy trend estimation model, whereas the model based on Gaussian observation noises is referred to as the Gaussian trend estimation model.

In this book, as a further improvement, we also treat the case of Gaussian observation noises with a time-varying variance. In this case, we replace the observation model (2.51) with the following model:

$$y_\lambda(n) = t_\lambda(n) + w_\lambda(n), \qquad w_\lambda(n) \sim N(0, \sigma_\lambda^2(n)/j(n)). \tag{2.52}$$

Note that $\sigma_\lambda^2(n)$ varies over time n and is estimated by the time-varying variance model reviewed in Sect. 2.1.2. We refer to this trend estimation model based on Gaussian observation noises with a time-varying variance (GTV) as the GTV trend estimation model.

Each extended trend model (2.49) with (2.51) or (2.52) can be expressed as a state-space model as follows:

$$x_\lambda(n) = F x_\lambda(n-1) + G v_\lambda(n) \tag{2.53}$$

$$y_\lambda(n) = H x_\lambda(n) + w_\lambda(n). \tag{2.54}$$

For example, for the trend order $k = 1$, the state vector is defined as $x_\lambda(n) = t_\lambda(n)$, and the matrices are defined as $F = G = H = 1$. For $k = 2$, $x_\lambda(n)$, F, G, and H are respectively defined as

$$x_\lambda(n) = \begin{bmatrix} t_\lambda(n) \\ t_\lambda(n-1) \end{bmatrix}, \quad F = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1, & 0 \end{bmatrix}.$$

Given a parameter λ of the Box–Cox transformation, the estimation of the state vector is performed in the following manner. When we assume that the observation noise distribution is Gaussian, i.e., either the Gaussian trend estimation model (2.51) or the GTV trend estimation model (2.52), the conditional means and conditional variance covariance matrices for the state can be calculated recursively by the Kalman filter. On the other hand, when the observation noise distribution is assumed to be non-Gaussian, e.g., a Cauchy distribution in (2.51), the state vector is estimated by a non-Gaussian filter or the sequential Monte Carlo filter presented in Sect. 2.1.4. In both cases, the parameters such as the variances of noises are estimated by the maximum likelihood method, and the missing observations, namely, observations that are not available for at certain points in time, can be interpolated by a smoothing algorithm (Anderson and Moore 2012; Kitagawa and Gersch 1996; Kitagawa 2010). The trend order k is selected using the AIC (Akaike 1998; Konishi and Kitagawa 2008). In addition, although the type of the observation noise distribution is entirely dependent on the observations, it can also be determined using the AIC.

One application in Chap. 4 compares Gaussian, Cauchy, and GTV observation noises using the data that was originally used in Tanokura et al. (2012).

This book proposes a trend model with Gaussian observation noises with a time-varying variance (GTV trend estimation model).

2.3 Construction of a Distribution-Free Index

The estimation method for the trend component $t_\lambda(n)$ in the trend model was presented in the previous section. Now, we have to search for an optimal λ in order to construct a distribution-free index for a financial market with non-Gaussian price distributions. Distribution-free means being impartial, regardless of the observation distributions.

For each λ, as mentioned in the previous section, assume that the average time series $y_\lambda(n),\ n = 1, \ldots, N$, in (2.48), of the Box–Cox transformed prices $q_{i,\lambda}(n)$ of the original prices $p_i(n),\ i = 1, \ldots, j(n)$, in (2.47), is modeled by the trend model (2.49) and (2.52).

Then, the one-step-ahead predictive density function of $y_\lambda(n)$ is given by

$$p(y_\lambda(n)|Y_{\lambda,n-1}) = \left\{ \frac{j(n)}{2\pi\sigma_\lambda^2(n)} \right\}^{\frac{1}{2}} \exp\left[-\frac{j(n)\{y_\lambda(n) - t_\lambda(n)\}^2}{2\sigma_\lambda^2(n)} \right], \tag{2.55}$$

where $Y_{\lambda,n-1} = \{y_\lambda(1), \ldots, y_\lambda(n-1)\}$. The log-likelihood and the AIC of the trend model for $y_\lambda(n)$ are respectively obtained as

$$\ell_\lambda = \sum_{n=1}^{N} \log p(y_\lambda(n)|Y_{\lambda,n-1}) \tag{2.56}$$

$$\begin{aligned} \mathrm{AIC}_\lambda &= -2\ell_\lambda + 2(\text{number of parameters}) \\ &= -2\ell_\lambda + 2(k+2). \end{aligned} \tag{2.57}$$

The optimal parameter λ should be determined with respect to the original prices. Therefore, the average time series $y_\lambda(n)$ is transformed back to $z_\lambda(n)$ by the following inverse Box–Cox transformation:

$$z_\lambda(n) = h_\lambda^{-1}(y_\lambda(n)) = \begin{cases} \{1 + \lambda\, y_\lambda(n)\}^{1/\lambda} & \lambda \neq 0 \\ \exp\ y_\lambda(n) & \lambda = 0. \end{cases} \tag{2.58}$$

Using the density function (2.55) of the trend model for $y_\lambda(n)$, the density function of the corresponding model for $z_\lambda(n)$ is given by

$$p(z_\lambda(n)|Z_{\lambda,n-1}) = \left| \frac{dh_\lambda}{dz} \right| p(y_\lambda(n)|Y_{\lambda,n-1}), \tag{2.59}$$

where $Z_{\lambda,n-1} = \{z_\lambda(1), \ldots, z_\lambda(n-1)\}$, and dh_λ/dz is the Jacobian of the Box–Cox transformation (2.47), which is obtained as $dh_\lambda/dz = z_\lambda(n)$. Then, the log-likelihood of the model evaluated on $z_\lambda(n),\ n = 1, \ldots, N$, is obtained as

$$\ell_\lambda^0 = \sum_{n=1}^{N} \log p(y_\lambda(n)|Y_{\lambda,n-1}) + \sum_{n=1}^{N} \log \left| \frac{dh_\lambda}{dz} \right|. \tag{2.60}$$

Moreover, AIC_λ is modified to AIC_λ^0, which is the AIC value of the corresponding model for $z_\lambda(n)$, and is evaluated as

$$\mathrm{AIC}_\lambda^0 = \mathrm{AIC}_\lambda - 2 \sum_{n=1}^{N} \log \left| \frac{dh_\lambda}{dz} \right|_{z=z_\lambda(n)} \tag{2.61}$$

(Kitagawa 2010).

Therefore, the optimal λ can be determined by minimizing AIC_λ^0 values, and the optimal trend component $t_\lambda(n),\ n = 1, \ldots, N$, in (2.52) is obtained.

Finally, returning to the original observations, the distribution-free index $i(n),\ n = 1, \ldots, N$, is defined as the inverse Box–Cox transformed values of the optimal trend component $t_\lambda(n)$, as follows:

$$i(n) = \begin{cases} \{1 + \lambda\, t_\lambda(n)\}^{1/\lambda} & \lambda \neq 0 \\ \exp\, t_\lambda(n) & \lambda = 0. \end{cases} \tag{2.62}$$

Let us observe the effect of using the Box–Cox transformation (2.47). Figure 2.8 shows the histograms of the Box–Cox transformed observations of the Japanese corporate CDS spreads illustrated in Fig. 2.7, where the optimal λ is given by -0.5. From back to front, the distribution at each point in time becomes closely symmetric and approximately Gaussian. When the mean is subtracted from the distribution at each point in time, the Box–Cox transformed distributions become easily understandable, as shown in Fig. 2.9. From back to front, the distribution at each point in time can approximately be regarded as Gaussian, even though the dispersion around the center largely varies over time. That is why the time-varying variance model is applied.

Note that the method for constructing a distribution-free index can be applied to general observations, such as the rate of return and the economic growth rate, which can be negative, although the Box–Cox transformation (2.47) is defined for observations taking positive values. Since the purpose of using the Box–Cox transformation is to search for an appropriate transformation close to a Gaussian distribution, the Box–Cox transformation can be applied to transformed observations in a positive domain by an appropriate distribution invariant function, say, a parallel transformation. In this way, we construct a distribution-free index for real GDP growth in Chap. 4.

In addition, since various transformations can be obtained by changing the parameter λ of the Box–Cox transformation, it is possible to examine the observation

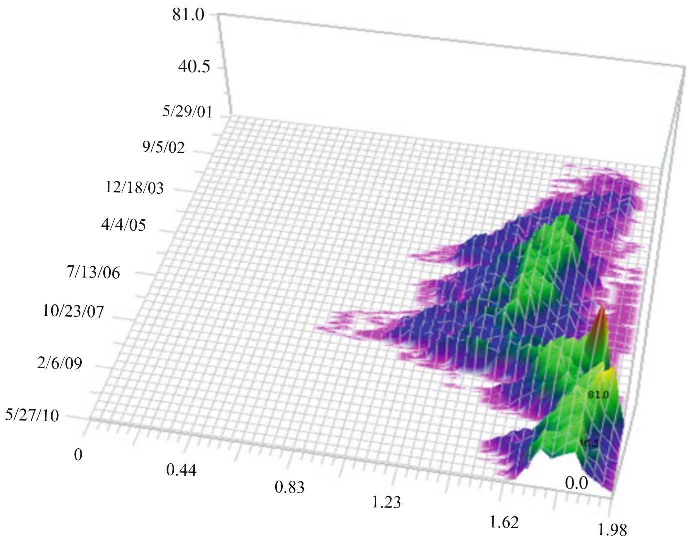


Fig. 2.8 Histograms of the Box–Cox transformed Japanese corporate CDS spreads, where $\lambda = -0.5$

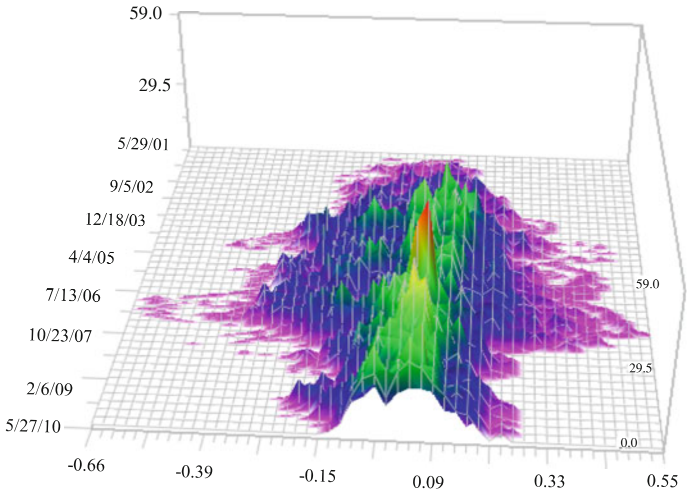


Fig. 2.9 Histograms of the mean subtracted at each point in time from the Box–Cox transformed Japanese corporate CDS spreads

distribution to be analyzed, namely, to determine how far the distribution is from a Gaussian distribution by $\mathrm{AIC}_{\lambda}^{0}$ values in (2.61). Moreover, this method can also be used to estimate the trend of a single time series.

We briefly describe a computation procedure for constructing a distribution-free index based on the trend model with Gaussian observation noises with a time-varying variance (GTV trend estimation model). The estimates of variances of τ_λ^2 and σ_λ^2 are obtained by the maximum likelihood method.

First, a trend order k is fixed.

1. Given a parameter $\lambda \in \{\lambda_1, \ldots, \lambda_m\}$ of the Box–Cox transformation.
   a. Transform the observations by the Box–Cox transformation (2.47).
   b. Estimate the parameters of the fitted trend model to the Box–Cox transformed observations by the Kalman filter.
   c. Compute the residual of the trend component.
   d. Estimate the time-varying variance by fitting the time-varying variance model in Sect. 2.1.2 to the residuals.
   e. Estimate the trend model with Gaussian observation noises with the above time-varying variance and compute AIC_λ^0 in (2.61).
2. Determine the optimal λ by minimizing AIC_λ^0.
3. Obtain the optimal trend model with the Gaussian observation noises with the time-varying variance.
4. The distribution-free index is obtained by the inverse Box–Cox transformation of the optimal trend component (2.62).

If necessary, the trend order k can be changed, and the above procedure can be repeated. As a reference, the standard trend estimation in Kitagawa (2010) can be helpful.

References

Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Parzen, E., Tanabe, K., Kitagawa, G. (eds.) Selected Papers of Hirotugu Akaike, pp. 199–213. Springer, New York (1998)

Anderson, B.D.O., Moore, J.B.: Optimal Filtering. Courier Dover Publications, Mineola (2012)

Box, G.E.P., Cox, D.R.: An analysis of transformations. J. Roy. Stat. Soc. Ser. B **26**(2), 211–252 (1964)

Davis, H.T., Jones, R.H.: Estimation of the innovations variance of a stationary time series. J. Am. Stat. Assoc. **63**, 141–149 (1968)

Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, New York (2001)

Fama, E.F.: The behavior of stock market prices. J. Bus. **38**(1), 34–105 (1965)

Gersch, W., Kitagawa, G.: The prediction of time series with trends and seasonalities. J. Bus. Econ. Stat. **1**(3), 253–264 (1983)

Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proc.-F **140**, 107–113 (1993)

Kitagawa, G.: Non-Gaussian state-space modeling of nonstationary time series. J. Am. Stat. Assoc. **82**(400), 1032–1041 (1987)

Kitagawa, G.: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J. Comput. Graph. Stat. **5**(1), 1–25 (1996)

Kitagawa, G.: Introduction to Time Series Modeling. CRC Press, Boca Raton (2010)

Kitagawa, G., Gersch, W.: A smoothness priors-state space modeling of time series with trend and seasonality. J. Am. Stat. Assoc. **79**(386), 378–389 (1984)

Kitagawa, G., Gersch, W.: Smoothness Priors Analysis of Time Series. Lecture Notes in Statistics, vol. 116. Springer, New York (1996)

Konishi, S., Kitagawa, G.: Information Criteria and Statistical Modeling. Springer, New York (2008)

Linden, M.: A model for stock return distribution. Int. J. Financ. Econ. **6**, 159–169 (2001)

Madan, D.B., Seneta, E.: The variance gamma (V. G.) model for share market returns. J. Bus. **63**(4), 511–524 (1990)

Mandelbrot, B.: The variation of certain speculative prices. J. Bus. **36**(4), 394–419 (1963)

Paraez, P.D.: The distribution of share price changes. J. Bus. **45**(4), 49–55 (1972)

Tanokura, Y., Tsuda, H., Sato, S., Kitagawa, G.: Constructing a credit default swap index and detecting the impact of the financial crisis. In: Bell, W.R., Holan, S.H., McElroy, T.S. (eds.) Economic Time Series: Modeling and Seasonality, pp. 359–380. CRC Press, Boca Raton (2012)

http://www.springer.com/978-4-431-55275-8

Indexation and Causation of Financial Markets
Tanokura, Y.; Kitagawa, G.
2015, X, 103 p. 50 illus., 33 illus. in color., Softcover
ISBN: 978-4-431-55275-8