Chapter 2

Theory and Methods

In this chapter we describe the data and summarise the statistical methods used to analyse them. These methods include standard methods for the analysis of time series data, as well as some newer methods for handling stochastic volatility, including some new theoretical developments.

Source of data

The original data comprise the exchange rates of the British pound sterling, the Japanese yen, and the German deutsche mark, relative to the US dollar. These rates change almost continuously in time. For convenience, observations are taken at the end of successive trading days, that is, days when the stock exchanges in these countries are open for business. As a result, the data do not satisfy the standard assumption made in time series analysis of equispaced observations in discrete time. However, since most economic movements tend to slow down during non-business days, for practical purposes the observations may be assumed to occur at discrete, equispaced, intervals of time.

The US dollar is chosen as the reference currency because the United States is the world's largest economy and its currency is thus the driving force in world financial markets.

The period selected for study is from 3 January 1986 to 12 April 1994, comprising 2158 successive trading days. This period is selected because it is comparatively recent and thus our findings are likely to be still relevant. But we have excluded the most recent period encompassing the Asian economic crisis, on the grounds that we do not wish our findings to be unduly influenced by the particular events associated with this phenomenon.

These data have been studied to some extent by other researchers, including Shephard (1996), who presented graphs of the distributions of the daily percentage returns for the yen and deutsche mark relative to the pound during this period. These graphs show that these returns have very long-tailed distributions, invalidating the basic statistical assumptions of normality.

Time series statistical methods

A time series is a set of numerical data measured sequentially in time. The measurements are often equispaced in time or nearly so. Time series data arise in many areas. These include economics and marketing (company sales or profits in successive months, stock market prices, currency relative values, etc.) and the physical sciences (barometric pressure in successive hours, snowfall in successive years, maximum and minimum daily air temperatures, etc.). Time series also arise in engineering processes (quality control charts, intervals between equipment failures, etc.), in biology and demography (sizes of animal population in successive seasons, birth and death rates in successive years, and university enrolments), and in many other applications areas.

The statistical analysis of time series data has shown the subject of many theoretical texts. These include Durbin (1996), Hamilton (1994), Diggle (1990), Tong (1990), Chatfield (1989), Abraham and Ledolter (1983), Whittle (1983), Anderson (1976), Bloomfeild (1976), Box and Jenkins (1976), Fuller (1976), Anderson, (1971), Hannan (1970), Jenkins and Watts (1968), Brown (1963), and Hannan (1960). However, as in many areas of statistics, modern developments in computer technology have made time series methods accessible to persons with a minimal background in mathematics.

Four important objectives arise in time series analysis. These are

1. Forecasting future values of a series.
2. Estimating the trend or overall character of a time series.
3. Modeling the dynamic relations between two or more time series.
4. Summarizing characteristic features of a time series.

1. Data transformation

Because time series methods are based on linear models, it is frequently necessary to transform the data. A logarithm transformation is usually needed for rates and financial data, whereas square roots are often better for transforming counts. The need for a transformation is usually apparent from an inspection of graph of the data. A more precise diagnosis, popularised by Tukey (1976), is obtained by grouping the data into relatively short intervals of time and graphing the standard deviation of these samples against their means, preferably using a logarithm scale on each axis. If a linear relation is apparent in this graph, the data should be transformed.

2. Removing a trend

Many time series have a trend. In these situations it may be useful to fit a straight line, or possibly a quadratic function, and use the residuals as a basis for further statistical analysis. Least squares regression may be used to fit a linear or quadratic trend to time series data.

3. Spectrum analysis

A time series is stationary if its statistical properties do not change with time. It is unlikely that a stationary time series will repeat itself exactly, but the series is repeatable in a probabilistic sense. Another way of looking at this is to say that the character of the series persists as you move forward or backward in time, and the only aspect that changes is the sampling error, which does not contain useful information. Of course these sampling fluctuations could be relatively large compared to the persistent characteristic. These ideas lead to the sinusoid (the simplest function that repeats itself) and to the idea of measuring the amount of periodicity or repeatability in a time series by finding its covariance or correlation with a sine wave having a given period. A sinusoid is characterized by the property that talking a linear transformation of its argument only shifts its frequency and its phase or position relative to some origin. The cosine function is just a sine function whose argument is shifted by $\pi/2$, that is

$$\cos(x) = \sin(x + \pi/2) \tag{1}$$

Since sinusoidal functions are periodic it is natural to use them as a basis for approximating a stationary time series. This basic comprises sine waves with different frequencies each defined on the time interval spanned by the data. The first component appears exactly once on this time interval, the second comprises two repeated sinusoids, the third three sinusoids, and so on. These components are also called *harmonics*. The functional form for the j^{th} harmonic is a cosine wave with some phase φ, that is, $\cos\{2\pi j(t-1)/n+\varphi\}$, $t = 1, 2, ..., n$. Using the mathematical theory of Fourier analysis any function defined at n equispaced points on a finite interval may be represented exactly by a constant plus $n-1$ harmonics. The number of different frequencies in these components, m, is $(n-1)/2$ or $n/2$ (depending on whether n is odd or even) since there is a sine and a cosine harmonic at each frequency. If n is even this Fourier representation takes the form

$$y_t = a_0+\sum[a_j\cos\{2\pi(t-1)/n\}+b_j\sin\{2\pi j(t-1)/n\}]+a_m\cos\{\pi(t-1)\} \qquad (2)$$

where the summation is from $j = 1$ to $j = m-1$. (Since $\sin\{\pi(t-1)\}$ is 0 for all integers t, in this case there is no sine harmonic at the highest frequency). A similar formula applies if n is odd. Using the fact that a linear combination of a sine function and a cosine function at the same frequency may be expressed as a single sinusoid with some phase φ, an alternative formula for the Fourier representation is

$$y_t = a_0 + \sum A_j\cos\{2\pi j(t-1)/n\}+\varphi_j \} \qquad (3)$$

where the amplitude $A_j = \sqrt{(a_j^2 + b_j^2)}$ and the summation is from 1 to m. This Fourier representation is similar to linear regression analysis, where the sinusoidal components play the role of determinants or predictor variables. Since the number of parameters is exactly equal to number of data values, there is no residual error: the regression model provides a perfect to the data. Moreover it may be shown that the sum of products or sine and/or cosine harmonics over the range of frequencies is zero, which means that these harmonics are statistically uncorrelated with each other. Consequently each Fourier coefficient (a_j or b_j) is the regression coefficient of the time series y_t on the corresponding harmonic. The formulas for these coefficients (for n even) are as follows:

$$a_0 = \sum y_t/n, \quad a_m = \sum(-1)^{t-1}y_t/n,$$
$$a_j = (2/n)\sum y_t\cos\{2\pi j(t-1)/n\},$$
$$b_j = (2/n)\sum y_t\sin\{2\pi j(t-1)/n\}.$$

We can see from these formulas that each Fourier coefficient may be interpreted as a covariance between the data and a sinusoid at the given frequency.

The *periodogram* of a time series (I_j, j = 1, 2, .., m) is defined in terms of the amplitudes of the harmonics in the Fourier representation as

$$I_j = (n/2)(a_j^2 + b_j^2) \tag{4}$$

The multiplier $n/2$ ensures that the j^{th} periodogram value is equal to the component of the variance in the data accounted for by sinusoidal function with frequency j/n. Since the sinusoidal terms are uncorrelated with each other, it follows that

$$\Sigma(y_t - \Sigma y_t/n)^2 = \Sigma I_j \tag{5}$$

This useful formula is known as *Parseval's theorem.* This relation is just an analysis of variance for a time series. So the sum of the periodogram ordinates is equal to the total squared error of the data, and consequently the periodogram shows how much of the squared error of the data is accounted for by the various harmonics. For this reason it useful to graph the scaled periodogram, obtained by dividing the periodogram by its sum. The scaled periodogram thus shows what proportion of the squared error is associated with each harmonic. Note that the frequency j/n is expressed in terms of the number of cycles per unit time. Since the values of j are 1, 2, ..., m, the lowest frequency is $1/n$, corresponding to a period equal to the whole range of the data, and the highest frequency is close to 0.5 (exactly 0.5 if n is even), corresponding to cycles of length 2 with the data oscillating from one value to the next.

4. Decomposition of a time series

A time series may be written in the form

$$y_t = p_t + s_t + z_t \tag{6}$$

where p_t is a trend (usually linear or quadratic), s_t is a stationary signal having the Fourier series representation given by Equation (5), and z_t is the residual, or *noise* series. In classical time series analysis, we assume that z_t has a normal distribution. In the simplest case, the terms in the process z_t are mutually uncorrelated, in which case the noise is called *white noise.*

Provided the noise is normally distributed, it may be shown that the periodogram coefficients are exponentially distributed. Now an exponential distribution has the property that its standard deviation is equal to its mean. However, the logarithm of an exponential distribution has approximately constant standard deviation. For this reason, it is useful, when analysing time series data, to plot the logarithm of the periodogram.

5. Testing for white noise

The periodogram and its logarithm may be used to investigate the character of a time series. Another useful graphical tool is the *correlogram,* or sample *autocorrelation function*, which comprises the set of estimated correlation coefficients between the series and itself at various spacings. Thus the (auto)correlation coefficient at spacing (or lag) s may be estimated from the formula

$$r_s = \frac{\sum_{t=1}^{n-s}(y_t - \bar{y})(y_{t+s} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{7}$$

and the correlogram is a graph of the series (r_s , $s = 1, 2, ..S$) against the spacing s. Since the number of terms used to calculate the correlation coefficient at lag s is $n-s$ where n is the length of the time series, the maximum spacing S should be substantially less than n. According to statistical theory, when the sample size n is large the standard error of a correlation coefficient is approximately normally distributed with standard deviation $1/\sqrt{n}$, which tends to 0 as n gets large. This means that as the length of an observed time series increases, the sample autocorrelation function of a stationary time series stabilizes, approaching a smooth curve. For a white noise process the theoretical correlation between observations at different spacing is zero, so you would expect the graph of its sample autocorrelation function to approach the horizontal axis $r = 0$ as n gets large.

Based on the normal distribution which has 95% of its probability within 1.96 standard deviations of its mean, a 95% confidence interval for the autocorrelation at lag s ranges from $-1.96/\sqrt{(n-s)}$ to $1.96/\sqrt{(n-s)}$. In contrast, the periodogram values of a white noise process, being exponentially distributed with constant standard deviation, do not settle down as the length of the series increases. Instead they become more densely packed, as we saw in the preceding section. Ljung & Box (1978) suggested using the statistic

$$Q = n(n+2)\sum_{s=1}^{m}\frac{r_s^2}{n-s} \tag{8}$$

where m is a specified integer substantially less than the series length n, to test the hypothesis that a time series is a sample from a white noise process. If it is necessary to fit a linear model involving p parameters to transform the series to a white noise process, where these parameters are estimated from the data, then Q is distributed approximately as a chi-squared distribution with $m-p$ degrees of freedom.

6. Autoregressive processes

Now let us consider more general models for describing a noise process z_t. A simple model, involving just a single parameter, takes the form

$$z_t = a_1 z_{t-1} + w_t \tag{9}$$

where w_t is a white noise process. This process is called a *simple Markov process*, and is characterised by the fact that the best forecast of its next value, z_{t+1}, is based only on the current value, z_t. Note that this process reduces to white noise when the parameter is 0.

This leads us to consider introducing a second parameter, extending the simple Markov process to the *second-order autoregressive* model, which takes the form

$$z_t = a_1 z_{t-1} + a_2 z_{t-2} + w_t \tag{10}$$

The general autoregressive process of order p takes the form

$$z_t = \Sigma\, a_j z_{t-j} + w_t \tag{11}$$

where the summation goes from $j=1$ to $j=p$.

It may be shown that an autoregressive process is stationary if and only if all of the roots of the *characteristic* polynomial

$$P(z) = 1 - \Sigma\, a_j z^j \tag{12}$$

are *outside* the unit circle $|z| = 1$ in the plane of complex numbers z. In particular, this means that a simple Markov process is stationary if $|a_1| < 1$. The condition for a second-order autoregressive process is rather more complicated, but it may be shown that necessary and sufficient conditions are

$$a_1 + a_2 < 1,\ \ a_2 - a_1 < 1,\ \ |a_2| < 1.$$

Stochastic volatility models

1. Moment generating function

The moment generating function of a random variable **X** is defined as the function

$$\Phi(\theta) = \mathrm{E}[\exp(-\theta \mathbf{X})] \tag{13}$$

where E denotes the *expected value* of a random variable and θ is a real number. Thus if **X** has probability density function $f(x)$, its moment generating function is

$$\Phi(\theta) = \int_{-\infty}^{\infty} \exp(-\theta x) f(x) dx \tag{14}$$

The moment generating function is useful because the moments of the distribution of X are obtained as mathematical derivatives of $\Phi(\theta)$, that is

$\mathrm{E}[X^n] = \Phi^{(n)}(\theta)(-1)^n$

This is because, using the Taylor series expansion of $\exp(-\theta x)$ about 0,

$$\Phi(\theta) = \mathrm{E}\left[1 - \theta x + \tfrac{1}{2}\theta^2 x^2 - \tfrac{1}{6}\theta^3 x^3 \ldots\right]$$

$$= 1 - \theta\, \mathrm{E}[X] + \tfrac{1}{2}\theta^2\, \mathrm{E}[X^2] - \tfrac{1}{6}\theta^3\, \mathrm{E}[X^3] + \ldots$$

If **X** has a normal (or Gaussian) distribution with mean μ and standard deviation σ, its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{15}$$

so its moment generating function is

$$\Phi(\theta) = \int_{-\infty}^{\infty} \exp(-\theta x) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx,$$

which may be evaluated to

$$\Phi(\theta) = \exp\left(-\theta\mu + \tfrac{1}{2}\theta^2\sigma^2\right). \tag{16}$$

2. Models for compounded returns

Suppose that x_t is the value of an asset on day t. The percentage compounded return is defined as

$$y_t = 100 \log_e\left(\frac{x_t}{x_{t-1}}\right) \tag{17}$$

In the simplest model, we assume that y_t is a process of independent, identically distributed random variables having a normal distribution with mean μ and standard deviation σ. In this case we can write

$$y_t = \mu + \sigma z_t \tag{18}$$

The parameter σ (or its square, in some definitions) is called the *volatility* of the compounded return.

Thus the logarithm of the asset price follows a random walk with drift, that is

$$\log x_t = \log x_{t-1} + \mu + \sigma z_t\,. \tag{19}$$

In this case can be shown that $\log x_t$ is normally distributed with mean μt and standard deviation $\sigma\sqrt{t}$. (see, for example, Taylor 1986)

Now consider the more general model

$$y_t = \mu + (\sigma + \delta u_t)\, z_t, \tag{20}$$

where u_t is a stationary first-order autoregressive process

$$u_t = \alpha u_{t-1} + w_t\,, \tag{21}$$

and w_t is Gaussian white noise (with mean 0 and standard deviation 1), and the correlation between z_t and w_{t+s} is $\rho\delta_s$ where $\delta_s = 1$ if $s = 0$, and 0 otherwise. Note that the condition for this process to be stationary is $|\alpha| < 1$. Since the volatility should be relatively smooth, we will assume that $\alpha \geq 0$. The volatility of y_t is thus $\sigma + \delta u_t$.

Since the volatility should be non-negative, the autoregressive process u_t should never be less than $-\sigma/\delta$, which means that there should be a reflecting barrier at this value. An alternative model, which ensures that the volatility is non-negative, has been suggested (see, for example, Shephard, 1996, page 22). This model may be written in the form

$$y_t = \mu + \sigma \exp(\delta u_t / \sigma)\, z_t \tag{22}$$

where u_t is again given by equation (21).

Models (20) and (22) may be incorporated into the more general model

$$y_t = \mu + \sigma \{1+\delta u_t/(\sigma k)\}^k z_t \tag{23}$$

When k is 1, this model reduces to (20), and (22) arises in the limit as k tends to infinity.

The following special cases of model (20) are of interest.

(a) When δ is 0, y_t is just Gaussian white noise, in which the volatility is constant.

(b) When σ is 0, the model corresponds to that considered by Heston (1993).

(c) When α is 0, the volatility itself is white noise, and is thus completely unpredictable.

(d) When $\delta \to 0$ and $\alpha \to 1$, in such a way that $\delta/\sqrt{(1-\alpha)}$ is constant, an interesting limiting process arises.

3. The stationary distribution of y_t

The model given by equations (20) and (21) has five parameters, μ, σ, δ, α, and ρ. To estimate them, we consider the stationary distribution of y_t .

First, we will derive the moment generating function of y_t, conditional on u_{t-1}. We proceed as follows. For simplicity, assume that μ is 0. If μ is not 0, we simply multiply the moment generating function by $\exp(-\theta\mu)$. Substituting equation (20),

$$\mathrm{E}[\,\exp(-\theta y_t) \mid u_{t-1}] = \mathrm{E}[\,\exp(-\theta(\sigma+\delta u_t)\, z_t) \mid u_{t-1}].$$

Substituting equation (21), this becomes

$$\mathrm{E}[\,\exp(-\theta y_t) \mid u_{t-1}] = \mathrm{E}[\,\exp(-\theta(\sigma+\delta\alpha\, u_{t-1} + \delta\, w_t)\, z_t)\,].$$

Since w_t and z_t are jointly distributed as a standardised bivariate normal distribution with correlation coefficient ρ, this expression may be written as

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\iint \exp\left(-\theta(\sigma+\delta\alpha u_{t-1}+\delta w)z-\frac{1}{2(1-\rho^2)}(w^2-2\rho wz+z^2)\right)dwdz.$$

The range of integration for both w and z is from $-\infty$ to $+\infty$. Completing the square on w and then integrating with respect to w, we obtain

$$\iint \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left\{w^2-2w\left(\rho-(1-\rho^2)\theta\delta\right)z+\left(\rho-(1-\rho^2)\theta\delta\right)^2 z^2\right\}\right)dw$$

$$\times\exp\left(\frac{1}{2(1-\rho^2)}\left\{\left(\rho-(1-\rho^2)\theta\delta\right)^2-1\right\}z^2-\theta(\sigma+\delta\alpha u_{t-1})z\right)dz\,,$$

$$=\int\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left\{1+2\rho\theta\delta-\theta^2\delta^2(1-\rho^2)\right\}\left\{z^2+2z\frac{\theta(\sigma+\delta\alpha u_{t-1})}{\kappa}+\frac{\theta^2(\sigma+\delta\alpha u_{t-1})^2}{\kappa^2}\right\}\right)dz$$

$$\times\exp\left(\frac{\theta^2(\sigma+\delta\alpha u_{t-1})^2}{2}\right),$$

where $\kappa = 1+2\rho\theta\delta-\theta^2\delta^2(1-\rho^2)$. Integrating one more time, this reduces to

$$\mathrm{E}[\exp(-\theta y_t)\mid u_{t-1}]=\frac{1}{\sqrt{1+2\delta\rho\theta-\delta^2(1-\rho^2)\theta^2}}\exp\left(\frac{\theta^2(\sigma+\delta\alpha u_{t-1})^2}{2(1+2\delta\rho\theta-\delta^2(1-\rho^2)\theta^2)}\right) \quad (24)$$

Finally, we integrate over the stationary distribution of u_{t-1}, which is normal with mean 0 and standard deviation $1/\surd(1-\alpha^2)$. This gives

$$\mathrm{E}[\exp(-\theta y_t)]=\frac{1}{\sqrt{\kappa}}\int\frac{\sqrt{1-\alpha^2}}{\sqrt{2\pi}}\exp\left(\frac{-\theta^2(\sigma+\delta\alpha u)^2}{2\kappa}-\frac{(1-\alpha^2)}{2}u^2\right)du$$

$$=\frac{1}{\sqrt{\kappa}}\int\frac{\sqrt{1-\alpha^2}}{\sqrt{2\pi}}\exp\left(-\frac{1-\alpha^2}{2}\left[u^2\left(1-\frac{\theta^2\delta^2\alpha^2}{(1-\alpha^2)\kappa}\right)-2u\left(\frac{\theta^2\delta\sigma\alpha\gamma}{(1-\alpha^2)\kappa\gamma}\right)\gamma+\left(\frac{\theta^2\delta\sigma\alpha}{(1-\alpha^2)\kappa\gamma}\right)^2\right]\right)$$

$$\times\exp\left(\frac{\theta^2\sigma^2}{2}+\frac{1-\alpha^2}{2}\left(\frac{\theta^2\delta\sigma\alpha}{(1-\alpha^2)\kappa\gamma}\right)^2\gamma\right)du$$

$$= \frac{1}{\sqrt{\kappa}}\frac{1}{\sqrt{\gamma}}\exp\left(\frac{\theta^2\sigma^2}{2}+\frac{1}{2}\left(\frac{\alpha^2}{1-\alpha^2}\right)\frac{\theta^4\delta^2\sigma^2}{\kappa^2\gamma}\right),$$

where $\gamma = 1-\theta^2\delta^2\alpha^2/\{\kappa(1-\alpha^2)\}$. Putting $\beta = \alpha/\sqrt{(1-\alpha^2)}$, and including the drift parameter μ, this expression finally simplifies to

E[exp($-\theta\, y$)]

$$= \frac{1}{\sqrt{1+2\rho\delta\theta-\delta^2(1-\rho^2+\beta^2)\theta^2}}\exp\left(-\mu\theta+\frac{\sigma^2\theta^2}{2\left(1+2\rho\delta\theta-\delta^2(1-\rho^2+\beta^2)\theta^2\right)}\right) \quad (25)$$

Note that when δ is 0, this function reduces to $\exp(-\mu\theta + \tfrac{1}{2}\sigma^2\theta^2)$, the moment generating function of a normal distribution with mean μ and standard deviation σ.

The moments may be obtained by expanding the moment generating function as a Taylor series about $\theta = 0$. The moments are then obtained by taking the coefficients of powers of θ. In this way, we obtain, are some lengthy but straightforward algebraic manipulations, the following expressions for the mean, standard deviation, skewness and kurtosis. Note that the skewness is defined as the third moment about the mean divided by the cube of the standard deviation, and the kurtosis is similarly defined as the fourth moment about the mean divided by the fourth power of the standard deviation.

$$mean[y] = \mu + \delta\rho \quad (26)$$

$$sd[y] = \sqrt{\sigma^2+\delta^2\left(1+\rho^2+\beta^2\right)} \quad (27)$$

$$sk[y] = \frac{2\delta\rho\left(3\sigma^2+\delta^2(3+\rho^2+3\beta^2)\right)}{\left(\sigma^2+\delta^2(1+\rho^2+\beta^2)\right)^{3/2}} \quad (28)$$

$$kur[y] = \frac{3\left(\sigma^4+2\sigma^2\delta^2(3+7\rho^2+3\beta^2)+\delta^4(3+14\rho^2+6\beta^2+6\rho^4+14\rho^2\beta^2+3\beta^4)\right)}{\left(\sigma^2+\delta^2(1+\rho^2+\beta^2)\right)^2} \quad (29)$$

These equations may be used to estimate the parameters, using the observed moments of the distribution of y. However, there are five parameters to be estimated, and only four moments available.

We can get further information by considering the joint distribution of y_t and y_{t-s}, for values of $s \geq 1$. It turns out that the correlation between y_t and y_{t-s} is 0. So let us derive a formula for the correlation between $(y_t)^2$ and $(y_{t-s})^2$, which is not zero. Using equations (20) and (21)

$$(y_t\, y_{t-s})^2 = (\sigma + \delta\alpha u_{t-1} + w_t)^2 z_t^2 (\sigma + \delta u_{t-s})^2 z_{t-s}^2$$

$$= \left(\sigma + \delta\alpha^2 u_{t-2} + \delta\sum_{j=0}^{s}\alpha^j w_{t-j}\right)^2 z_t^2 (\sigma + \delta\alpha u_{t-s-1} + \delta w_{t-s}) z_{t-s}^2 .$$

Taking the expected value of this expression, using the facts that

(a) w_t and z_t are jointly distributed as a standardised bivariate normal distribution with correlation coefficient ρ, and

(b) the stationary distribution of u_{t-s} is normal with mean 0 and standard deviation β/α, we obtain, again after some lengthy but straightforward algebraic calculations, the following result.

$Corr[(y_t)^2,(y_{t-s})^2]$

$$= \frac{\delta^2\left(2\alpha^s\sigma^2(1+2\rho^2+\beta^2)+\alpha^{2s}\{\rho^2\sigma^2+\delta^2(1+5\rho^2+2\beta^2+5\rho^2\beta^2+\beta^4)\}\right)}{\sigma^4+2\sigma^2\delta^2\left(4+17\rho^2+4\beta^2\right)+2\delta^4\left(2+17\rho^2+4\beta^2+5\rho^4+17\rho^2\beta^2+2\beta^4\right)} \tag{30}$$

The estimation problem is still quite difficult.

For simplicity, let us assume for the moment that the model for the process y_t is given by the special limiting case (d), that is, $\delta \to 0$ and $\alpha \to 1$, in such a way that $\delta\beta = \kappa$, a constant.

In this case, with $s = 1$, equations (26) – (30) reduce to the following:

$$mean[y] = \mu \tag{31}$$

$$sd[y] = \sqrt{\sigma^2 + \kappa^2} \tag{32}$$

$$sk[y] = 0 \tag{33}$$

$$kur[y] = \frac{3\left(\sigma^4 + 6\kappa^2\sigma^2 + 3\kappa^4\right)}{\left(\sigma^2 + \kappa^2\right)^2} \tag{34}$$

$$Corr[(y_t)^2,(y_{t-s})^2] = \frac{\alpha^s\kappa^2(2\sigma^2 + \alpha^s\kappa^2)}{\sigma^4 + 8\sigma^2\kappa^2 + 4\kappa^4} \tag{35}$$

The estimates of the parameters are thus

$$\hat{\mu} = mean(y) \tag{36}$$

$$\hat{\sigma}^2 = s^2 \sqrt{(1.5 - kurtosis/6)} \tag{37}$$

$$\hat{\kappa} = \sqrt{(s^2 - \hat{\sigma}^2)} \tag{38}$$

Equation (35) may now be used to see how well the model fits thc data.

In the general case, equations (26) – (30) may be solved iteratively.