LECTURE 1

Trends in Economic Time Series

In many time series, broad movements can be discerned which evolve more gradually than the other motions which are evident. These gradual changes are described as trends and cycles. The changes which are of a transitory nature are described as fluctuations.

In some cases, the trend should be regarded as nothing more than the accumulated effect of the fluctuations. In other cases, we feel that the trends and the fluctuations represent different sorts of influences, and we are inclined to decompose the time series into the corresponding components.

In economics, it is traditional to decompose time series into a variety of components, some or all of which may be present in a particular instance. If $\{Y_t\}$ is the sequence of values of an economic index, then its generic element is liable to be expressed as

$$Y_t = T_t + C_t + S_t + \varepsilon_t, \tag{1.1}$$

where

T_t is the global trend,

C_t is a secular cycle,

S_t is the seasonal variation and

ε_t is an irregular component.

Many of the more prominent macroeconomic indicators are amenable to a decomposition of the sort depicted above. One can imagine, for example, a quarterly index of Gross National Product which appears to be following an exponential growth trend $\{T_t\}$.

The growth trend might be obscured, to some extent, by a superimposed cycle $\{C_t\}$ with a period of roughly four and a half years, which happens to correspond, more or less, to the average lifetime of the legislative assembly. The reasons for this curious coincidence need not concern us here.

The ghost of an annual cycle $\{S_t\}$ might also be apparent in the index; and this could well be a reflection of the fact that some economic activities,

such as building construction, are significantly affected by the weather and by the duration of sunlight.

When the foregoing components—the trend, the secular cycle and the seasonal cycle—have been extracted from the index, the residue should correspond to an irregular component $\{\varepsilon_t\}$ for which no unique explanation can be offered. This component ought to resemble a time series generated by a so-called stationary stochastic process. Such a series has the characteristic that any segment of consecutive elements looks much like any other segment of the same duration, regardless of the date at which it begins or ends.

If the residue follows a trend, or if it manifests a more or less regular pattern, then it contains features which ought to have been attributed to the other components; and we should set about the task of redefining them.

There are two distinct purposes for which we might wish to effect such a decomposition. The first purpose is to give a summary description of the salient features of the time series. Thus, if we eliminate the irregular and seasonal components from the series, we are left with an index which may give a clearer picture of the more important features. This might help us to gain an insight into the fundamental workings of the economic or social structure which has generated the series.

The other purpose in decomposing the series is to predict its future values. For each component of the time series, a particular method of prediction is appropriate. By combining the separate predictions of the components, a forecast can be derived which may be superior to one derived by a method which pays no attention to the underlying structure of the time series.

Extracting the Trend

There are essentially two ways of extracting trends from a time series. The first way is to apply to the series a variety of so-called filters which annihilate or nullify all of the components which are not regarded as trends.

A filter is a carefully crafted moving average which spans a number of data points and which attributes a weight to each of them. The weights should sum to unity to ensure that the filter does not systematically inflate or deflate the values of the series. Thus, for example, the following moving average might serve to eliminate the annual cycle from an economic series which is recorded at quarterly intervals:

$$\hat{Y}_t = \frac{1}{16}\Big\{Y_{t+3} + 2Y_{t+2} + 3Y_{t+1} + 4Y_t + 3Y_{t-1} + 2Y_{t-2} + Y_{t-3}\Big\}. \tag{1.2}$$

Another filter with a wider span and a different profile of weights might serve to eliminate the four-and-a-half-year cycle which is present in our imaginary series of Gross National Product.

Finally a filter could be designed which smooths away the irregularities of the index which defy systematic explanation. The order in which the three filters are applied is immaterial; and what is left after they have been applied should give a picture of the underlying trend $\{T_t\}$ of the index.

Other collections of filters, applied in series, might serve to isolate the other components $\{C_t\}$ and $\{S_t\}$ which are to be found in equation (1).

The process of filtering is often a good way of deriving an index which represents the more important historical characteristics of the time series. However, it generates no model for the underlying trends; and it suggests no way of predicting their future values.

The alternative way of extracting the trend from the index is to fit some function which is capable of adapting itself to whatever form the trend happens to display. Different functions are appropriate to different forms of trend; and some functions which analysts tend to favour see almost always to be inappropriate. Once an analytic function has been fitted to the series, it may be used to provide extrapolative forecasts of the trend.

Polynomial Trends

Amongst the mathematical functions which suggest themselves as means of modelling a trend is a pth-degree polynomial whose argument is the time index t:

$$\phi(t) = \phi_0 + \phi_1 t + \cdots + \phi_p t^p. \tag{1.3}$$

When there is no theory to specify a mathematical form for the trend, it may be possible to approximate it by a polynomial of low degree. This notion is suggested by the formal result that every analytic mathematical function can be expanded as a power series, which is an indefinite sum whose terms contain rising powers of the argument. Thus the polynomial in t may be construed as an approximation to an analytic function which is obtained by discarding all but the leading terms of a power-series expansion.

There are also arguments from physics which suggest that first-degree and second-degree polynomials in t, which are linear and quadratic time trends in other words, are common in the natural world. The thought occurs to us that such trends might also arise in the social world.

According to a well-known dictum,

> *Every body continues in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed upon it.*

This is Newtons's first law of motion. The kinematic equation for the distance covered by a body moving with constant velocity in a straight line is

$$x = x_0 + ut, \tag{1.4}$$

where u is the uniform velocity, and x_0 represents the initial position of the body at time $t = 0$. This is nothing but a first-degree polynomial in t.

Newton's second law of motion asserts that

> *The change of motion is proportional to the motive force impressed; and is made in the direction of the straight line in which the force is impressed.*

In modern language, this is expressed by saying that the acceleration of a body along a straight line is proportional to the force which is applied in that direction. The kinematic equation for the distance travelled under uniformly accelerated rectilinear motion is

$$x = x_0 + u_0 t + \frac{1}{2}at^2, \tag{1.5}$$

where u_0 is the velocity at time $t = 0$ and a is the constant acceleration due to the motive force. This is just a quadratic in t.

A linear or a quadratic function may be appropriate if the trend in question is monotonically increasing or decreasing. In other cases, polynomials of higher degrees might be fitted. Figure 1 is the result of fitting a cubic function to an economic time series by least-squares regression.

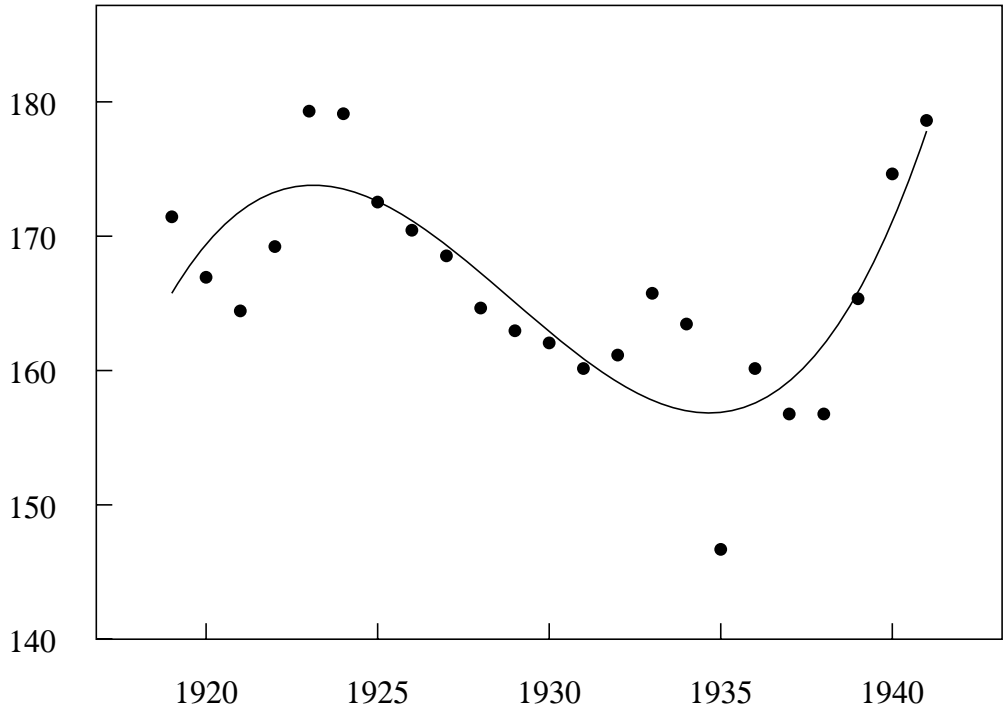

Figure 1. A cubic function fitted to data on meat consumption in the United States, 1919–1941.

It might be felt that there are salient features in the data which are not captured by the cubic polynomial. In that case, the recourse might be to increase the degree of the polynomial by one. The result will be a curve which fits the data more closely. Also, it will be found that one of the branches of the polynomial—the left branch in this case—has changed direction. The values found by extrapolating the quartic function backwards in time will differ radically from those found by extrapolating the cubic function.

In general, the effect of altering the degree of the polynomial by one will be to alter the direction of one or other of the branches of the fitted function; and, from the point of view of forecasting, this is a highly unsatisfactory circumstance. Another feature of a polynomial function is that its branches tend to plus or minus infinity with increasing rapidity as the argument increases or decreases beyond a range of central values where the function has its stationary points and its points of inflection. This might also be regarded as a undesirable property for a function which is to be used in extrapolative forecasting.

Some care has to be taken in fitting a polynomial time trend by the method of least-squares regression. A straightforward procedure, which comes immediately to mind, is to form a matrix X of regressors in which the generic row $[t^0, t, t^2, \ldots, t^p]$ contains rising powers of the argument t. The annual data on meat consumption, for example, which are plotted in Figure 1, run from 1919 to 1941; and these dates might be taken as the initial and terminal values of t. In that case, there would be a vast differences in the values of the elements of the matrix X. For, whereas $t^0 = 1$ for all values of $t = 1919, \ldots, 1941$, we should find that, when $t = 1941$, the value of t^3 is in excess of $7,300$ million. Clearly, such a disparity of numbers taxes the precision of the computer.

An obvious recourse is to recode the values of t. Thus, we might take $t = -11, \ldots, 11$ for the range of the argument. The change would affect only the value of the intercept term ϕ_0 which could be adjusted *ex post*. Unfortunately, such a recourse in not always adequate to ensure the numerical accuracy of the computation. The reason lies in the peculiarly ill-conditioned nature of the matrix $(X'X)^{-1}$ of cross products.

In fact, a specialised procedure of polynomial regression is often called for in which the functions $t^0, t, \ldots, t^p$ are replaced by a set of so-called orthogonal polynomials which give rise to vectors of regressors whose cross products are zero-valued. The estimated coefficients associated with these orthogonal polynomials can be converted into the coefficients $\phi_0, \phi_1, \ldots, \phi_p$ of equation (3).

Exponential and Logistic Trends

The notion of exponential or geometric growth is common in economics where it is closely related to the idea of compound interest. Consider a financial asset with an annual rate of return of γ. The annual growth factor for an

investment of unit value is $(1 + \gamma)$. If α units were invested at time $t = 0$, and if the returns were compounded with the principal on an annual basis, then the value of the investment at time t would be given by

$$y_t = \alpha(1 + \gamma)^t. \tag{1.6}$$

An investment which is compounded twice a year has an annual growth factor of $(1+\frac{1}{2}\gamma)^2$, and one which is compounded quarterly has a growth factor of $(1 + \frac{1}{4}\gamma)^4$. If an investment were compounded continuously, then its growth factor would be $\lim(n \to \infty)(1 + \frac{1}{n}\gamma)^n = e^\gamma$. The value of the asset at time t would be given by

$$y = \alpha e^{\gamma t}; \tag{1.7}$$

and this is the equation for exponential growth.

The equation of exponential growth is a solution of the differential equation

$$\frac{dy}{dt} = \gamma y. \tag{1.8}$$

The implication of the differential equation is that the absolute rate of growth in y is proportional to the value already attained by y. It is equivalent to say that the proportional rate of growth $(1/y)(dy/dt)$ is constant.

An exponential growth trend can be fitted to observations $y_1, \ldots, y_n$, sampled at regular intervals, by applying ordinary least-squares regression to the equation

$$\ln y_t = \ln \alpha + \gamma t + \varepsilon_t. \tag{1.9}$$

This is obtained by taking the logarithm of equation (7) and adding a disturbance term ε_t. An alternative parametrisation is obtained by setting $\lambda = e^\gamma$. Then the transformed growth equation becomes

$$\ln y_t = \ln \alpha + (\ln \lambda)t + \varepsilon_t, \tag{1.10}$$

and the geometric growth rate is $\lambda - 1$.

Whereas unhindered exponential growth might well be a possibility for certain monetary or financial quantities, it is implausible to suggest that such a process can be sustained for long when real resources are involved. Since real resources are finite, we expect there to be upper limits to the levels which can be attained by economic variables.

For an example of a trend with an upper bound, we might imagine a process whereby the ownership of a consumer durable grows until the majority

of households or individuals are in possession of it. Good examples are provided by the sales of domestic electrical appliances such are fridges and colour television sets.

Typically, when the new durable is introduced, the rate of sales is slow. Then, as information about the durable, or experience of it, is spread amongst consumers, the sales begin to accelerate. For a time, their cumulated total might appear to follow an exponential growth path. Then come the first signs that the market is being saturated; and there is a point of inflection in the cumulative curve where its second derivative—which is the rate of increase in sales per period—passes from positive to negative. Eventually, as the level of ownership approaches the saturation point, the rate of sales will decline to a constant level, which may be at zero, if the good is wholly durable, or at a small positive replacement rate if it is not.

It is very difficult to specify the dynamics of a process such as the one we have described whenever there are replacement sales to be taken into account. The reason is that the replacement sales depend not only on the size of the ownership of the durable goods but also upon the age of the stock of goods. The latter is a function, at least in an early period, of the way in which sales have grown at the outset. Often we have to be content with modelling only the growth of ownership.

One of the simplest ways of modelling the growth of ownership is to employ the so-called logistic curve. This classical device has its origins in the mathematics of biology where it has been used to model the growth of a population of animals in an environment with limited food resources.

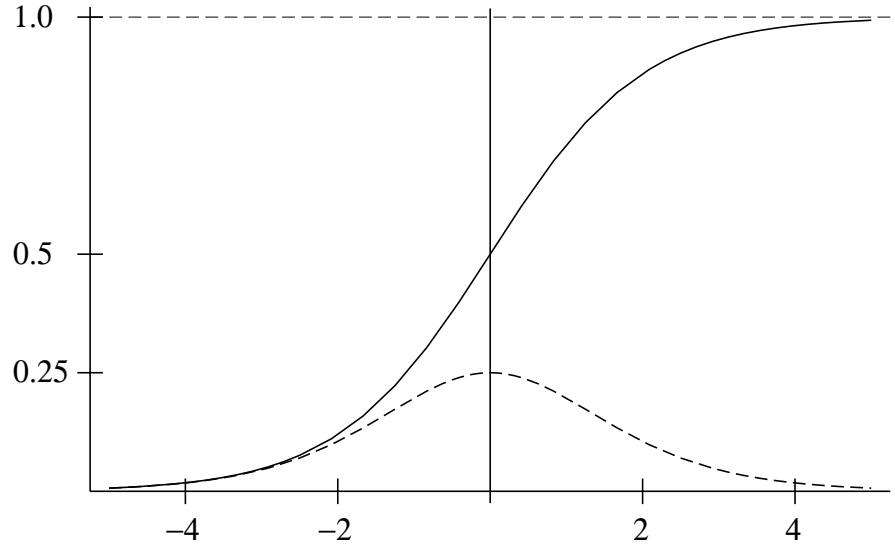

Figure 2. The logistic function $e^x/(1+e^x)$ and its derivative. For large negative values of x, the function and its derivative are close. In the case of the exponential function e^x, they coincide for all values of x.

The simplest version of the function is given by

$$\pi(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}. \tag{1.11}$$

The second expression comes from multiplying top and bottom of the first expression by e^x. The logistic curve varies between a value of zero, which is approached as $x \to -\infty$, and a value of unity, which is approached as $x \to +\infty$. At the mid point, where $x = 0$, the value of the function is $\pi(0) = \frac{1}{2}$. These characteristics can be understood easily in reference to the first expression.

The alternative expression for the logistic curve also lends itself to an interpretation. We may begin by noting that, for large negative values of x, the term $1+e^x$, which is found in the denominator, is not significantly different from unity. Therefore, as x increases from such values towards zero, the logistic function closely resembles an exponential function. By the time x reaches zero, the denominator, with a value of 2, is already significantly affected by the term e^x. At that point, there is an inflection in the curve as the rate of increase in π begins to decline. Thereafter, the rate of increase declines rapidly toward zero, with the effect that the value of π never exceeds unity.

The inverse mapping $x = x(\pi)$ is easily derived. Consider

$$\begin{aligned} 1 - \pi &= \frac{1 + e^x}{1 + e^x} - \frac{e^x}{1 + e^x} \\ &= \frac{1}{1 + e^x} = \frac{\pi}{e^x}. \end{aligned} \tag{1.12}$$

This is rearranged to give

$$e^x = \frac{\pi}{1 - \pi}, \tag{1.13}$$

whence the inverse function is found by taking natural logarithms:

$$x(\pi) = \ln\left\{\frac{\pi}{1 - \pi}\right\}. \tag{1.14}$$

The logistic curve needs to be elaborated before it can be fitted flexibly to a set of observations $y_1, \ldots, y_n$ tending to an upper asymptote. The general from of the function is

$$y(t) = \frac{\gamma}{1 + e^{-h(t)}} = \frac{\gamma e^{h(t)}}{1 + e^{h(t)}}; \quad h(t) = \alpha + \beta t. \tag{1.15}$$

Here γ is the upper asymptote of the function, which is the saturation level of ownership in the example of the consumer durable. The parameters β and α

determine respectively the rate of ascent of the function and the mid point of its ascent, measured on the time-axis.

It can be seen that

$$\ln\left\{\frac{y(t)}{\gamma - y(t)}\right\} = h(t). \tag{1.16}$$

Therefore, with the inclusion of a residual term, the equation for the generic element of the sample is

$$\ln\left\{\frac{y_t}{\gamma - y_t}\right\} = \alpha + \beta t + e_t. \tag{1.17}$$

For a given value of γ, one may calculate the value of the dependent variable on the LHS. Then the values of α and β may be found by least-squares regression.

The value of γ may also be determined according to the criterion of minimising the sum of squares of the residuals. A crude procedure would entail running numerous regressions, each with a different value for γ. The definitive value would be the one from the regression with the least residual sum of squares. There are other procedures for finding the minimising value of γ of a more systematic and efficient nature which might be used instead. Amongst these are the methods of Golden Section Search and Fibonnaci Search which are presented in many texts of numerical analysis.

The objection may be raised that the domain of the logistic function is the entire real line—which spans all of time from creation to eternity—whereas the sales history of a consumer durable dates only from the time when it is introduced to the market. The problem might be overcome by replacing the time variable t in equation (15) by its logarithm and by allowing t to take only nonnegative values. Then, whilst $t \in [0, \infty)$, we still have $\ln(t) \in (-\infty, \infty)$, which is the entire domain of the logistic function.

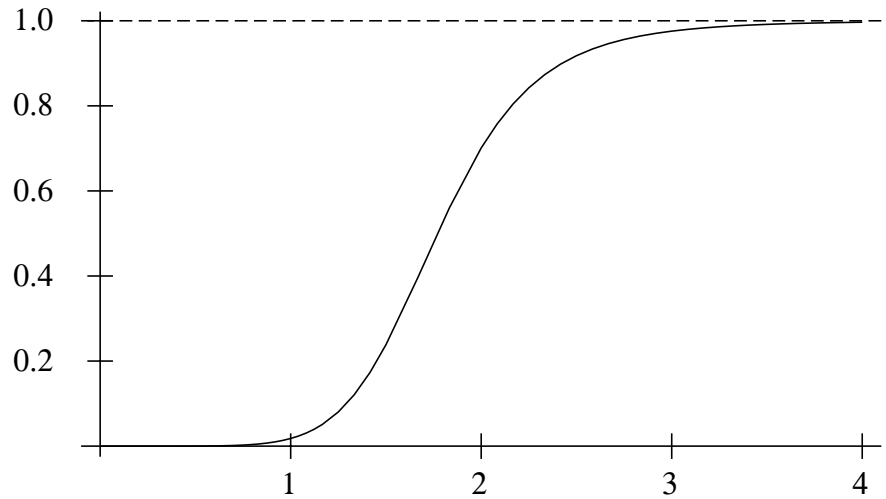

Figure 3. The function $y(t) = \gamma/(1 + \exp\{\alpha - \beta \ln(t)\})$ with $\gamma = 1$, $\alpha = 4$ and $\beta = 7$. The positive values of t are the domain of the function.

There are many curves which will serve the purpose of modelling a sigmoidal growth process. Their number is equal, at least, to the number of theoretical probability density functions—for the corresponding (cumulative) distribution functions rise monotonically from zero to unity in ways with are suggestive of processes of bounded growth.

In fact, we do not need to have an analytic form for a cumulative function before it can be fitted to a growth process. It is enough to have a table of values of a standardised form of the function. An example is provided by the normal density function whose distribution function is regularly fitted to data points in the course of probit analysis. In this case, the fitting involves finding values for the location parameter μ and the dispersion parameter σ^2 by which the standard normal function is converted into an arbitrary normal function. Nowadays, there are efficient procedures for numerical optimisation which can accomplish such tasks with ease.

Flexible Trends

If the purpose of decomposing a time series is to form predictions of its components, then it is important to obtain adequate representations of these components at every point within the sample period. The device which is most appropriate to the extrapolative forecasting of a trend is rarely the best means of representing it within the sample. An extrapolation is usually based upon a simple analytic function; and any attempt to make the function reflect the local variations of the sample will endow it with global characteristics which may affect the forecasts adversely.

One way of modelling the local characteristics of a trend without prejudicing its global characteristics is to use a segmented curve. In many applications, it has been found that a curve with cubic polynomial segments is appropriate. The segments must be joined in a way which avoids evident discontinuities. In practice, the requirement is usually for continuous first-order and second-order derivatives. A curve whose segments are joined in this way is described as a cubic spline.

A spline is a draughtsman's tool which was once used in drawing smooth curves. It is a thin flexible piece of wood which was clamped to a series of pins which were placed along the path of the curve which had to be described. Some of the essential properties of a mathematical spline can be understood by bearing the real spline in mind. The pins to which a draughtsman clamped his spline correspond to the data points through which we might interpolate a mathematical spline. The segments of the mathematical spline would be joined at the data points.

The cubic spline becomes a device for modelling a trend when, instead of passing through the data points, it is allowed, in the interests of smoothness, to deviate from them. The Reinsch smoothing spline is fitted by minimising

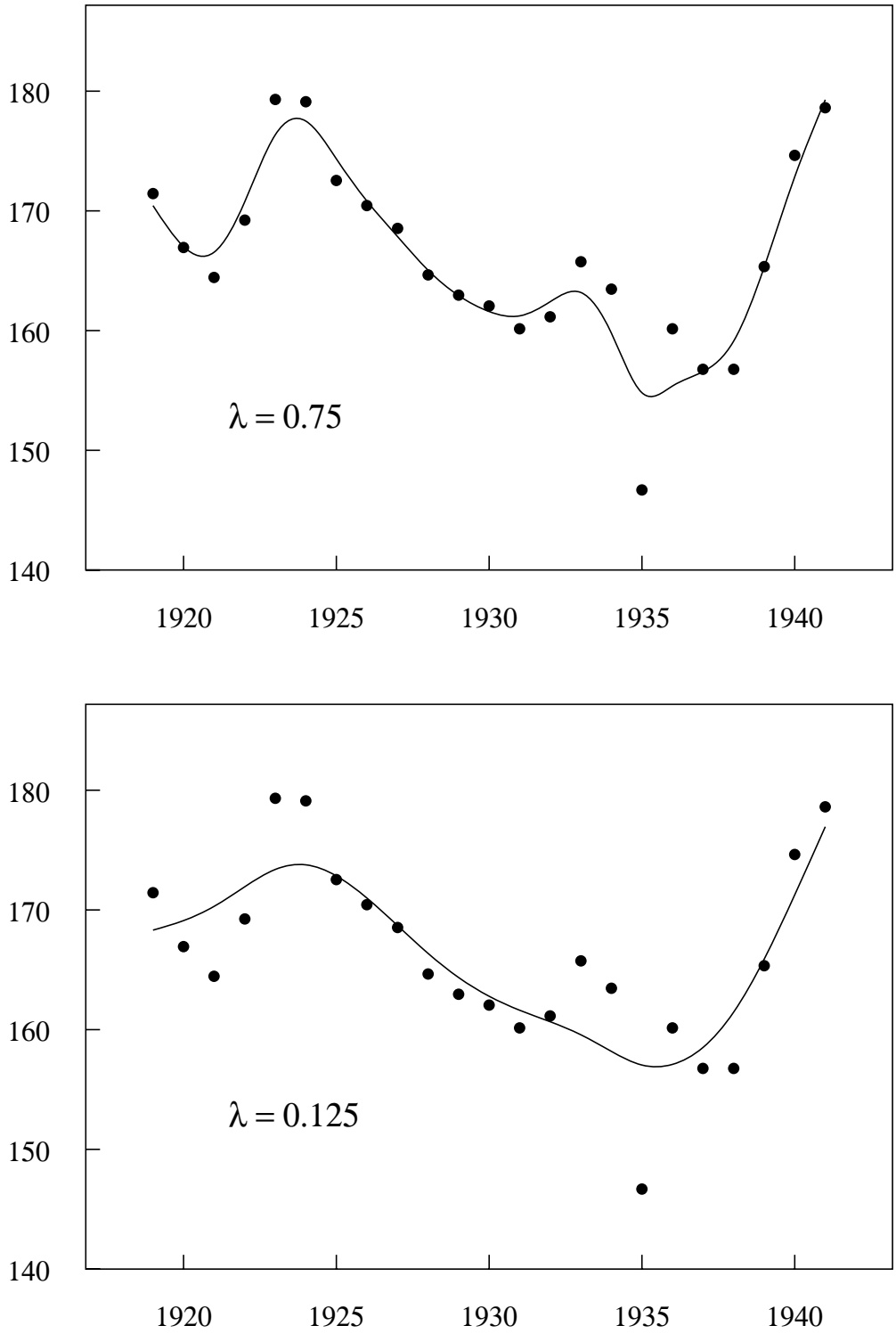

Figure 4. Cubic smoothing splines fitted to data on meat consumption in the United States, 1919–1941.

a criterion function which imposes both a penalty for deviating from the data points and a penalty for excessive curvature in the segments. The measure of curvature is based upon second derivatives, whilst the measure of deviation is the sum of the squared distances of the points from the curve. A single parameter λ governs the trade-off between the objectives of smoothness and goodness of fit.

As an analogy for the smoothing spline, one might think of attaching the draughtsman's spline to the pins by springs instead of by clamps. The precise form of the curve would depend upon the stiffness of the spline and the forces exerted by the springs. The degree of flexibility of the spline corresponds to the value of λ. The forces exerted by ordinary springs are proportional to their extension; and, in this respect, the analogy, which requires the forces to be proportional to the squares of their extensions, is imperfect.

Figure 4 shows the consequences of fitting the smoothing spline to the data on meat consumption which is also used in Figure 1 where a cubic polynomial has been fitted. It is a matter of judgment how the value of λ should be chosen so as to reflect the trend.

There are various ways in which the curve of a cubic spline may be extrapolated to form forecasts of the trend. In normal circumstances, when the ends of the spline are left free, the second derivatives are zero-valued and the extrapolation is linear. However, it is possible to clamp the ends of the spline in a way which imposes a value on their first derivatives. In that case, the extrapolation is quadratic.

Stochastic Trends

It is possible that what is perceived as a trend is the result of the accumulation of small stochastic fluctuations which have no systematic basis. In that case, there are some clearly defined ways of removing the trend from the data as well as for extrapolating it into the future.

The simplest model embodying a stochastic trend is the so-called first-order random walk. Let $\{y_t\}$ be the random-walk sequence. Then its value at time t is obtained from the previous value via the equation

$$y_t = y_{t-1} + \varepsilon_t. \tag{1.18}$$

Here ε_t is an element of a white-noise sequence of independently and identically distributed random variables with

$$E(\varepsilon_t) = 0 \quad \text{and} \quad V(\varepsilon_t) = \sigma^2 \quad \text{for all } t. \tag{1.19}$$

By a process of back-substitution, the following expression can be derived:

$$y_t = y_0 + \big\{\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1\big\}. \tag{1.20}$$

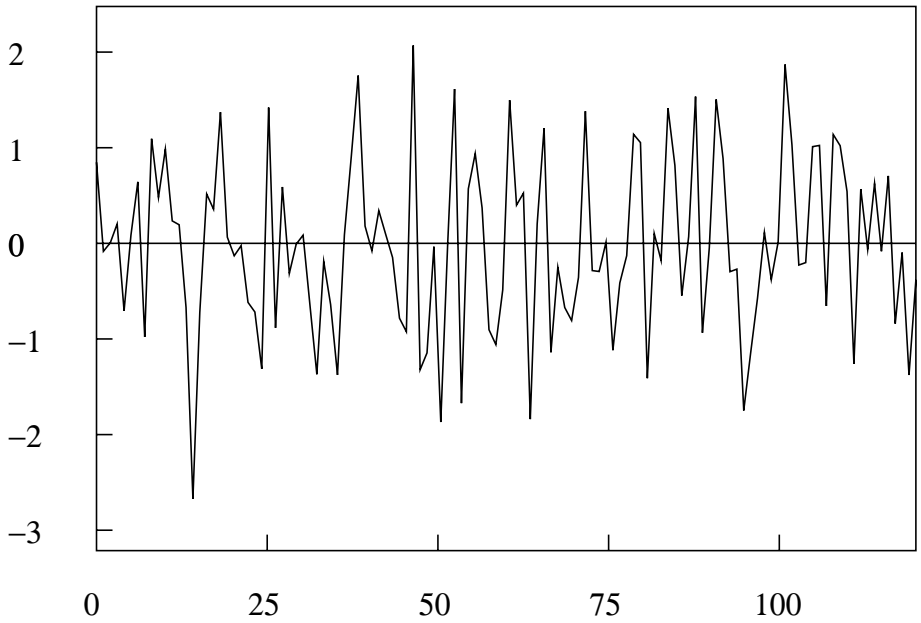

Figure 5. A sequence generated by a white-noise process.

This depicts y_t as the sum of an initial value y_0 and of an accumulation of stochastic increments. If y_0 has a fixed finite value, then the mean and the variance of y_t are be given by

$$E(y_t) = y_0 \quad \text{and} \quad V(y_t) = t \times \sigma^2. \tag{1.21}$$

There is no central tendency in the random-walk process; and, if its starting point is in the indefinite past rather than at time $t = 0$, then the mean and variance are undefined.

To reduce the random walk to a stationary stochastic process, it is necessary only to take its first differences. Thus

$$y_t - y_{t-1} = \varepsilon_t. \tag{1.22}$$

The values of a random walk, as the name implies, have a tendency to wander haphazardly. However, if the variance of the white-noise process is small, then the values of the stochastic increments will also be small and the random walk will wander slowly. It is debatable whether the outcome of such a process deserves to be called a trend.

A first-order random walk over a surface is what is know as Brownian motion. For a physical example of Brownian motion, one can imagine small particles, such a pollen grains, floating on the surface of a viscous liquid. The viscosity might be expected to bring the particles to a halt quickly if they

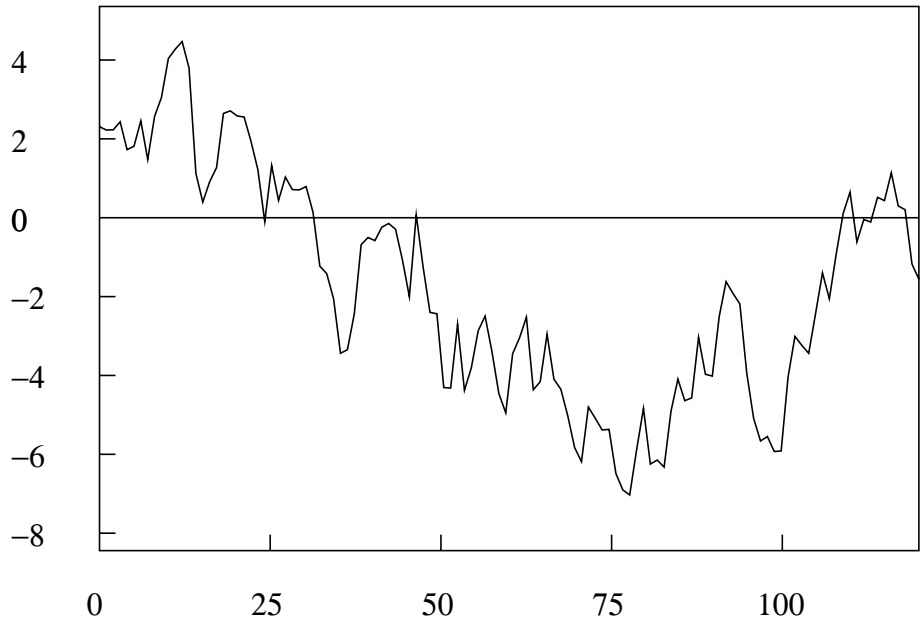

Figure 6. A first-order random walk

Figure 7. A second-order random walk

were in motion. However, if the particles are very light, then they will dart hither and thither on the surface of the liquid under the impact of its molecules which are themselves in constant motion.

There is no better way of predicting the outcome of a random walk than to take the most recently observed value and to extrapolate it indefinitely into the future. This is demonstrated by taking the expected values of the elements of the equation

$$y_{t+h} = y_{t+h-1} + \varepsilon_{t+h} \tag{1.23}$$

which represents the value which lies h periods ahead at time t. The expectations, which are conditional upon the information of the set $\mathcal{I}_t = \{y_t, y_{t-1}, \ldots\}$ containing observations on the series up to time t, may be denoted as follows:

$$E(y_{t+h}|\mathcal{I}_t) = \begin{cases} \hat{y}_{t+h|t}, & \text{if } h > 0; \\ y_{t+h}, & \text{if } h \leq 0. \end{cases} \tag{1.24}$$

In these terms, the predictions of the values of the random walk for $h > 1$ periods ahead and for one period ahead are given, respectively, by

$$\begin{aligned} E(y_{t+h}|\mathcal{I}_t) &= \hat{y}_{t+h|t} = \hat{y}_{t+h-1|t}, \\ E(y_{t+1}|\mathcal{I}_t) &= \hat{y}_{t+1|t} = y_t. \end{aligned} \tag{1.25}$$

The first of these, which comes from (23), depends upon the fact that $E(\varepsilon_{t+h}|\mathcal{I}_t) = 0$. The second, which comes from taking expectations in the equation $y_{t+1} = y_t + \varepsilon_{t+1}$, uses the fact that the value of y_t is already known. The implication of the two equations is that y_t serves as the optimal predictor for all future values of the random walk.

A second-order random walk is formed by accumulating the values of a first-order process. Thus, if $\{\varepsilon_t\}$ and $\{y_t\}$ are respectively a white-noise sequence and the sequence from a first-order random walk, then

$$\begin{aligned} z_t &= z_{t-1} + y_t \\ &= z_{t-1} + y_{t-1} + \varepsilon_t \\ &= 2z_{t-1} - z_{t-2} + \varepsilon_t \end{aligned} \tag{1.26}$$

defines the second-order random walk. Here the final expression is obtained by setting $y_{t-1} = z_{t-1} - z_{t-2}$ in the second expression. It is clear that, to reduce the sequence $\{z_t\}$ to the stationary white-noise sequence, we must take first differences twice in succession.

The nature of a second-order process can be understood by recognising that it represents a trend in which the slope—which is its first difference—follows a random walk. If the random walk wanders slowly, then the slope of

this trend is liable to change only gradually. Therefore, for extended periods, the second-order random walk may appear to follow a linear time trend.

For a physical analogy of a second-order random walk, we can imagine a body in motion which suffers a series of small impacts. If the kinetic energy of the body is large relative to the energy of the impacts, then its linear motion will be disturbed only slightly. In order to predict where the body might be in some future period, we simply extrapolate its linear motion free from disturbances.

To demonstrate that the forecast function for a second-order random walk is a straight line, we may take the expectations, which are conditional upon $\mathcal{I}_t$, of the elements of the the equation

$$z_{t+h} = 2z_{t+h-1} - z_{t+h-2} + \varepsilon_{t+h}. \tag{1.27}$$

For h periods ahead and for one period ahead, this gives

$$\begin{aligned} E(z_{t+h}|\mathcal{I}_t) &= \hat{z}_{t+h|t} = 2\hat{z}_{t+h-1|t} - \hat{z}_{t+h-2|t}, \\ E(z_{t+1}|\mathcal{I}_t) &= \hat{z}_{t+1|t} = 2z_t - z_{t-1}, \end{aligned} \tag{1.28}$$

which together serve to define a simple iterative scheme. It is straightforward to confirm that these difference equations have an analytic solution of the form

$$\hat{z}_{t+h|t} = \alpha + \beta h \quad \text{with} \quad \alpha = z_t \quad \text{and} \quad \beta = z_t - z_{t-1}, \tag{1.29}$$

which generates a linear time trend.

It is possible to define random walks of higher orders. Thus a third-order random walk is formed by accumulating the values of a second-order process. A third-order process can be expected to give rise to local quadratic trends; and the appropriate way of predicting its values is by quadratic extrapolation.

A stochastic trend of the random-walk variety may be elaborated by the addition of an irregular component. A simple model consists of a first-order random walk with an added white-noise component. The model is specified by the equations

$$\begin{aligned} y_t &= \xi_t + \eta_t, \\ \xi_t &= \xi_{t-1} + \nu_t, \end{aligned} \tag{1.30}$$

wherein η_t and ν_t are generated by two mutually independent white-noise processes.

The equations combine to give

$$\begin{aligned} y_t - y_{t-1} &= \xi_t - \xi_{-1} + \eta_t - \eta_{t-1} \\ &= \nu_t + \eta_t - \eta_{t-1}. \end{aligned} \tag{1.31}$$

The expression on the RHS can be reformulated to give

$$\nu_t + \eta_t - \eta_{t-1} = \varepsilon_t - \mu\varepsilon_{t-1}, \tag{1.32}$$

where ε_t and ε_{t-1} are elements of a white-noise sequence and μ is a parameter of an appropriate value. Thus, the combination of the random walk and white noise gives rise to the single equation

$$y_t = y_{t-1} + \varepsilon_t - \mu\varepsilon_{t-1}. \tag{1.33}$$

The forecast for h steps ahead, which is obtained by taking expectations in the equation $y_{t+h} = y_{t+h-1} + \varepsilon_{t+h} - \mu\varepsilon_{t+h-1}$, is given by

$$E(y_{t+h}|\mathcal{I}_t) = \hat{y}_{t+h|t} = \hat{y}_{t+h-1|t}. \tag{1.34}$$

The forecast for one step ahead, which is obtained from the equation $y_{t+1} = y_t + \varepsilon_{t+1} - \mu\varepsilon_t$, is

$$\begin{aligned} E(y_{t+1}|\mathcal{I}_t) = \hat{y}_{t+1|t} &= y_t - \mu\varepsilon_t \\ &= y_t - \mu(y_t - \hat{y}_{t|t-1}) \\ &= (1-\mu)y_t + \mu\hat{y}_{t|t-1}. \end{aligned} \tag{1.35}$$

The equation $\hat{y}_{t+1|t} = y_t - \mu\varepsilon_t$ reflects the fact that, if the information at time t consists of the elements of the set $\mathcal{I}_t = \{y_t, y_{t-1}, \ldots\}$ and the value of μ, then ε_t is a know quantity which is unaffected by the process of taking expectations. Taking the equation back on period gives $\hat{y}_{t|t-1} = y_{t-1} - \mu\varepsilon_{t-1}$, and subtracting the latter from (33) yields the identity $\varepsilon_t = y_t - \hat{y}_{t|t-1}$ upon which the second equality of (35) depends.

By applying a straightforward process of back-substitution to the final equation of (35), it will be found that

$$\begin{aligned} \hat{y}_{t+1|t} &= (1-\mu)(y_t + \mu y_{t-1} + \cdots + \mu^{t-1}y_1) + \mu^t\hat{y}_0 \\ &= (1-\mu)\{y_t + \mu y_{t-1} + \mu^2 y_{t-2} + \cdots\}, \end{aligned} \tag{1.36}$$

where the final expression stands for an infinite series. This is a so-called exponentially-weighted moving average; and it is the basis of the widely-used forecasting procedure known as exponential smoothing.

To form the one-step-ahead forecast $\hat{y}_{t+1|t}$ in the manner indicated by the first of the equations under (36), an initial value $\hat{y}_0$ is required. Equation (34) indicates that all the succeeding forecasts $\hat{y}_{t+2|t}, \hat{y}_{t+3|t}$ etc. have the same value as the one-step-ahead forecast.

It will transpire, in subsequent lectures, that equation (33) is a simple example of an Integrated Autoregressive Moving-Average or ARIMA model. There exists a readily accessible general theory of the forecasting of ARIMA processes which we shall expound at length.

References

Eubank, R.L., (1988), Spline Smoothing and Nonparametric Regression, Marcel Dekker Inc. New York.

Hamming, R.W., (1989), Digital Filters: Third Edition, Prentice–Hall Inc., Englewood Cliffs, N.J.

Ratkowsky, D.L., (1985), Nonlinear Regression Modelling: A Unified Approach, Marcel Dekker Inc. New York.

Reinsch, C.H., (1967), "Smoothing by Spline Functions", Numerische Mathematik, 10, 177–183.

Schoenberg, I.J., (1964), "Spline Functions and the Problem of Graduation", Proceedings of the National Academy of Science, 52, 947–950.

De Vos, A.F. and I.J. Steyn, (1990), "Stochastic Nonlinearity: A Firm Basis for the Flexible Functional Form": Research Memorandum 1990-13, Vrije Universiteit Amsterdam.