A Simple Method of Trend Construction

By C. E. V. Leser
School of General Studies, Australian National University

[Received June 1960]

Summary

The principle adopted here in the construction of a trend for a time series consists in minimizing a linear combination of two sums of squares, of which one refers to the second differences of the trend values, the other to the deviations of the observations from the trend values. Properties of the general solution are deduced, and the solution is explicitly obtained for up to 7 observations.

In the special case in which the sum of the two sums of squares is minimized, the exact solution is derived for up to 15 observations. An approximation formula, suitable for practical use when there are 8 or more observations, is also given. The method is illustrated by examples, in which it is applied to an artificially constructed and to an actual time series.

1. Introduction

The two textbook methods of fitting a trend to a time series, by means of moving averages or fitted regression curves, are well known. The drawbacks of either method are also familiar. When fitting a regression curve, for example, the straight line is often found to be insufficiently flexible. This difficulty may be overcome by the use of curvilinear regression; but whilst the introduction of the second and higher powers of time into the equation may seem natural in biometric work, there is often little justification for applying the procedure to economic data. Furthermore, the higher the degree of the fitted polynomial, the greater is usually the danger of getting misleading results when using the trend for extrapolation. It is also necessary to recalculate the entire trend given by a regression line or curve, once a new observation has been added to the data.

The moving average is therefore often preferred. It has been shown, for example by Macaulay (1931) and Kendall (1948), that moving averages with particular sets of weights can in fact be considered as moving regression curves. A moving average, of course, does not lend itself to prediction; but more than that, it does not yield trend values for the beginning and the end of the period under consideration. This loss is particularly serious in short time series.

Whittaker (1923, 1924) developed a method, also reproduced by Whittaker and Robinson (1926), in which he extended the principle of least squares, to make a linear combination of two sums of squares a minimum. The terms which are squared in the first sum are the third differences of successive trend values, those in the second sum the deviations of the observations from the trend. The method has been further developed by Henderson (1924) and is therefore known in this form as the Whittaker–Henderson method.

Although the possibility of using the sum of the squares of the second, instead of the third, differences was mentioned by Macaulay (1931), it does not seem to have been followed up. Yet this procedure seems a particularly natural one when dealing with economic time series. The resulting family of trends may be described as quasi-linear trends. One particular member of the family, which may be called the central quasi-linear trend, specially recommends itself on theoretical grounds, is easy to work out, and an approximate solution valid for any number of observations may be written down by means of a simple formula.

In the present study, particular emphasis is placed upon short time series, for which ordinary moving-average methods are clearly inadequate, and on presenting the solution in a form in which it readily lends itself to practical application.

2. The General Method

We take as our starting-point a time series consisting of n successive observations at unit intervals of time ($n \geqslant 3$). For the sake of argument, the time interval may be taken as a year; if there is no seasonal variation inherent in the data, the time interval could be shorter, say a quarter or a month.

A theoretical basis for the method to be developed is furnished by the assumption that the time series would tend to follow a straight line if unaffected by disturbances; where this assumption is unrealistic, a logarithmic or other transformation is first applied to the original data. The disturbances are thought to be partly of a permanent, partly of a temporary, nature; the permanent disturbances become incorporated in the trend as changes in direction, whilst the temporary disturbances leave the trend unaffected. The permanent disturbances are defined as the second differences of successive trend values, the temporary disturbances as deviations of the observations from their respective trend values. With n observations we have n temporary but only $n-2$ permanent disturbances, suitably allocated to the observations excluding the first and last one.

These relations may be written as follows:

$$\left.\begin{aligned} \gamma_i &= \theta_{i-1} - 2\theta_i + \theta_{i+1} \quad && (i = 2, 3, \ldots, n-1) \\ \epsilon_i &= Y_i - \theta_i && (i = 1, 2, \ldots, n), \end{aligned}\right\} \tag{1}$$

where Y_i represents the ith observation, θ_i the unknown corresponding trend value, γ_i the permanent, and ϵ_i the temporary disturbances.

Time may also be introduced as a variable, and the n successive points of time denoted by X_i ($i = 1, 2, \ldots, n$) where $X_{i+1} - X_i = 1$ ($i = 1, 2, \ldots, n-1$).

An alternative possibility would have been to define the permanent disturbances as

$$\theta_i - \tfrac{1}{2}(\theta_{i-1} + \theta_{i+1}) = -\tfrac{1}{2}\gamma_i;$$

that is to say, the difference between the trend value and the value which would have been obtained by linear interpolation between the preceding and the following trend value.

To choose the trend, a natural extension of the least squares principle suggests taking the sum of squares for the permanent disturbances and the sum of squares for the temporary disturbances, and setting the condition that a linear combination of the two sums be minimized, that is to say, minimizing

$$\sum_{i=2}^{n-1} \gamma_i^2 + h \sum_{i=1}^{n} \epsilon_i^2 \quad (h \geqslant 0). \tag{2}$$

The same condition can be derived from maximum likelihood principles. Suppose for the moment that not only all values of the time series, but also all but one of the trend values were observable, and the trend value θ_i corresponding to time X_i only had to be estimated. Assume further that the permanent and temporary disturbances are independently and normally distributed, and that their population variances are in the ratio $h:1$. Then the exponent of the likelihood function for obtaining the neighbouring trend values θ_{i-2}, θ_{i-1}, θ_{i+1} and θ_{i+2} (or, if $i<3$ or $i>n-2$, those that are defined) and the actual value Y_i will be, apart from a constant factor, $\gamma_{i-1}^2+\gamma_i^2+\gamma_{i+1}^2+h\epsilon_i^2$ (or those terms that are defined). All other values of γ_j^2 and ϵ_j^2 can be treated as constants in this context and can be added. Thus if (2) is applied to θ_i alone, it yields a maximum likelihood estimate for θ_i, given all θ_j $(j \neq i)$. The same consideration can successively be applied to all θ_i.

According to (1),

$$\frac{\partial \gamma_i}{\partial \theta_j} = \begin{matrix} 1 & (j=i-1, i+1), \\ -2 & (j=i), \\ 0 & (j \neq i-1, i, i+1), \end{matrix}$$

$$\frac{\partial \epsilon_i}{\partial \theta_j} = \begin{matrix} -1 & (j=i), \\ 0 & (j \neq i). \end{matrix}$$

Thus successive partial differentiation of (2) with regard to $\theta_1, \theta_2, \ldots, \theta_n$ leads to n linear equations which can be conveniently written in the form

$$g_{i-1}-2g_i+g_{i+1} = he_i \quad (i=1, \ldots, n), \tag{3}$$

if we denote the estimates of γ_i and ϵ_i by g_i and e_i and adopt the convention that

$$g_0 = g_1 = g_n = g_{n+1} = 0.$$

Equations (3) can be rewritten so as to obtain n linear equations for the trend values in terms of the data. According to whether $n=3$ or $n \geqslant 4$, we have, writing T_i for the estimate of θ_i,

$$\begin{aligned} n=3: \quad (1+h)T_1-2T_2+T_3 &= hY_1, \\ -2T_1+(4+h)T_2-2T_3 &= hY_2, \\ T_1-2T_2+(1+h)T_3 &= hY_3; \end{aligned}$$

$$\begin{aligned} n \geqslant 4: \quad (1+h)T_1-2T_2+T_3 &= hY_1, \\ -2T_1+(5+h)T_2-4T_3+T_4 &= hY_2, \\ T_1-4T_2+(6+h)T_3-4T_4+T_5 &= hY_3, \\ &\cdots\cdots\cdots \\ T_{n-4}-4T_{n-3}+(6+h)T_{n-2}-4T_{n-1}+T_n &= hY_{n-2}, \\ T_{n-3}-4T_{n-2}+(5+h)T_{n-1}-2T_n &= hY_{n-1}, \\ T_{n-2}-2T_{n-1}+(1+h)T_n &= hY_n. \end{aligned} \tag{4}$$

It is understood that for $n=4$, only the first two and the last two equations are operative.

7

Thus, given n and h, the trend values are linear functions of the data and can be written as

$$T_i = \sum_{j=1}^{n} K_{i,j} Y_j \quad (i = 1, \ldots, n). \tag{5}$$

Before proceeding to give numerical values for the coefficients $K_{i,j}$ in selected cases, some theoretical properties of the solution may be derived.

In the first place, since the matrix in (4) is doubly symmetric, the same applies to its inverse, and

$$K_{i,j} = K_{j,i} = K_{n+1-i,\, n+1-j} = K_{n+1-j,\, n+1-i}. \tag{6}$$

There are thus only $\frac{1}{4}(n+1)^2$ different coefficients if n is odd, and $\frac{1}{4}n(n+2)$ different coefficients if n is even.

Secondly, by adding all equations of (4) we find that

$$\sum_{i=1}^{n} T_i = \sum_{j=1}^{n} Y_j.$$

In conjunction with (5) this gives

$$\sum_{i=1}^{n} K_{i,j} = 1 \quad (j = 1, \ldots, n)$$

and thus

$$\sum_{j=1}^{n} K_{i,j} = 1 \quad (i = 1, \ldots, n). \tag{7}$$

The trend thus represents a weighted arithmetic mean of the observations, provided the definition of the mean is so stretched as to admit negative weights. It will be seen that negative coefficients do in fact occur.

Furthermore, by multiplying equations (4) successively by the consecutive integers $X_1, X_2, \ldots, X_n$ respectively and adding, we arrive at

$$\sum_{i=1}^{n} X_i T_i = \sum_{j=1}^{n} X_j Y_j$$

$$\sum_{i=1}^{n} X_i K_{i,j} = X_j \quad (j = 1, \ldots, n),$$

$$\sum_{j=1}^{n} X_j K_{i,j} = X_i \quad (i = 1, \ldots, n). \tag{8}$$

It also follows immediately that

$$\sum_{i=1}^{n} e_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} X_i e_i = 0;$$

that is to say, the temporary disturbances as estimated here have zero mean and are uncorrelated with time. This result could also have been directly deduced from (3). No simple condition of this kind exists for the permanent disturbances; these may, for example, all have the same sign. This implies a high degree of flexibility in the shape of the trend.

If all observations lie on a straight line, i.e.

$$Y_i = a + bX_i \quad (i = 1, \ldots, n),$$

then according to (5), (7) and (8),

$$T_i = a \sum_{j=1}^{n} K_{i,j} + b \sum_{j=1}^{n} X_j K_{i,j} = a + bX_i;$$

in other words, the trend coincides with that straight line. This, of course, is no more than one should expect.

Suppose now that the true trend is a straight line

$$\eta_i = \alpha + \beta X_i \quad (i = 1, \ldots, n)$$

and that the observations contain in addition an error term with variance δ^2, i.e.

$$Y_i = \alpha + \beta X_i + \delta_i \quad (i = 1, \ldots, n),$$

with $$E(\delta_i) = 0, \qquad E(\delta_i \delta_j) = 0 \quad (j \neq i), \qquad E(\delta_i^2) = \delta^2.$$

Then $$T_i = \alpha + \beta X_i + \sum_{j=1}^{n} K_{i,j} \delta_j,$$

$$E(T_i) = \alpha + \beta X_i = \eta_i,$$

$$E(T_i - \eta_i)^2 = \delta^2 \sum_{j=1}^{n} K_{i,j}^2.$$

Thus T_i will give an unbiased estimate of η_i, though the accuracy of the estimate does not necessarily increase with n as in the case of linear regression.

The effect on the solution of varying h may be considered next. It is immediately seen from (4) that when $h \to \infty$, $T_i \to Y_i$. If we put $h = 0$, any set of trend values T_i lying on a straight line will satisfy equations (4); but if we let h tend towards 0, it can be shown that the solution tends towards the elementary regression line. In the present terminology, we can write as $h \to 0$:

$$T_1 \to \frac{2}{n^3 - n}\{(2n^2 - 3n + 1) Y_1 + (2n^2 - 6n + 4) Y_2 + \ldots + (-n^2 + 3n - 2) Y_n\},$$

$$T_2 \to \frac{2}{n^3 - n}\{(2n^2 - 6n + 4) Y_1 + (2n^2 - 9n + 13) Y_2 + \ldots + (-n^2 + 6n - 5) Y_n\},$$

....

$$T_n \to \frac{2}{n^3 - n}\{(-n^2 + 3n - 2) Y_1 + (-n^2 + 6n - 5) Y_2 + \ldots + (2n^2 - 3n + 1) Y_n\}.$$

For other values of h, the solution of (4) cannot readily be written in general form. By successive elimination, (4) has been solved for $n = 3$, 4, 5, 6, and 7. Since for a given n, all coefficients $K_{i,j}$ are obtained as ratios with a common denominator D, Table 1 gives D and $DK_{i,j}$.

TABLE 1

General solution of equations (4) *in form* (5) *for* $n \leqslant 7$

$n=3$:					
D	$6+$	h			
$DK_{1,1}$	$5+$	h			
$DK_{1,2}$	2				
$DK_{1,3}$	-1				
$DK_{2,2}$	$2+$	h			
$n=4$:					
D	$20+$	$12h+$	h^2		
$DK_{1,1}$	$14+$	$11h+$	h^2		
$DK_{1,2}$	$8+$	$2h$			
$DK_{1,3}$	$2-$	h			
$DK_{1,4}$	-4				
$DK_{2,2}$	$6+$	$7h+$	h^2		
$DK_{2,3}$	$4+$	$4h$			
$n=5$:					
D	$50+$	$75h+$	$18h^2+$	h^3	
$DK_{1,1}$	$30+$	$63h+$	$17h^2+$	h^3	
$DK_{1,2}$	$20+$	$20h+$	$2h^2$		
$DK_{1,3}$	$10-$	$3h-$	h^2		
$DK_{1,4}$	$-$	$6h$			
$DK_{1,5}$	$-10+$	h			
$DK_{2,2}$	$15+$	$31h+$	$13h^2+$	h^3	
$DK_{2,3}$	$10+$	$22h+$	$4h^2$		
$DK_{2,4}$	$5+$	$8h-$	h^2		
$DK_{3,3}$	$10+$	$37h+$	$12h^2+$	h^3	
$n=6$:					
D	$105+$	$328h+$	$166h^2+$	$24h^3+$	h^4
$DK_{1,1}$	$55+$	$253h+$	$148h^2+$	$23h^3+$	h^4
$DK_{1,2}$	$40+$	$106h+$	$32h^2+$	$2h^3$	
$DK_{1,3}$	$25+$	$9h-$	$9h^2-$	h^3	
$DK_{1,4}$	$10-$	$28h-$	$6h^2$		
$DK_{1,5}$	$-5-$	$20h+$	h^2		
$DK_{1,6}$	$-20+$	$8h$			
$DK_{2,2}$	$31+$	$113h+$	$92h^2+$	$19h^3+$	h^4
$DK_{2,3}$	$22+$	$80h+$	$46h^2+$	$4h^3$	
$DK_{2,4}$	$13+$	$41h+$	$3h^2-$	h^3	
$DK_{2,5}$	$4+$	$8h-$	$8h^2$		
$DK_{3,3}$	$19+$	$126h+$	$92h^2+$	$18h^3+$	h^4
$DK_{3,4}$	$16+$	$100h+$	$40h^2+$	$4h^3$	

TABLE 1 (*continued*)

General solution of equations (4) *in form* (5) *for* $n \leqslant 7$

$n = 7$:	
D	$196+1{,}134h+1{,}050h^2+293h^3+30h^4+h^5$
$DK_{1,1}$	$91+\ 806h+\ 884h^2+269h^3+29h^4+h^5$
$DK_{1,2}$	$70+\ 400h+\ 264h^2+\ 44h^3+\ 2h^4$
$DK_{1,3}$	$49+\ 99h-\ 28h^2-\ 15h^3-\ h^4$
$DK_{1,4}$	$28-\ 62h-\ 64h^2-\ 6h^3$
$DK_{1,5}$	$7-\ 97h-\ 15h^2+\ h^3$
$DK_{1,6}$	$-14-\ 48h+\ 10h^2$
$DK_{1,7}$	$-35+\ 36h-\ h^2$
$DK_{2,2}$	$56+\ 358h+\ 476h^2+189h^3+25h^4+h^5$
$DK_{2,3}$	$42+\ 246h+\ 288h^2+\ 70h^3+\ 4h^4$
$DK_{2,4}$	$28+\ 134h+\ 76h^2-\ 3h^3-\ h^4$
$DK_{2,5}$	$14+\ 50h-\ 30h^2-\ 8h^3$
$DK_{2,6}$	$-\ 6h-\ 34h^2+\ h^3$
$DK_{3,3}$	$35+\ 344h+\ 505h^2+183h^3+24h^4+h^5$
$DK_{3,4}$	$28+\ 302h+\ 278h^2+\ 64h^3+\ 4h^4$
$DK_{3,5}$	$21+\ 190h+\ 52h^2-\ 2h^3-\ h^4$
$DK_{4,4}$	$28+\ 386h+\ 470h^2+183h^3+24h^4+h^5$

In addition to $h = 0$ and $h = \infty$, two particular values of h suggest themselves: the value $h = 1$ which minimizes

$$\sum_{i=2}^{n-1} \gamma_i^2 + \sum_{i=1}^{n} \epsilon_i^2,$$

and the value $h = 4$ which minimizes

$$\sum_{i=2}^{n-1} (\tfrac{1}{2}\gamma_i)^2 + \sum_{i=1}^{n} \epsilon_i^2;$$

that is to say, they minimize the sum of squares of all disturbances by either the definition adopted or the alternative definition which would be possible. The values assumed by the coefficients $K_{i,j}$ in these cases are given in Table 2 for $n \leqslant 7$.

TABLE 2

Solution for particular values of h, $n \leqslant 7$

		$h = 0$	$h = 1$	$h = 4$	$h = \infty$
$n = 3$:	$K_{1,1}$	·8333	·8571	·9	1
	$K_{1,2}$	·3333	·2857	·2	0
	$K_{1,3}$	−·1667	−·1429	−·1	0
	$K_{2,2}$	·3333	·4286	·6	1

TABLE 2 (*continued*)

Solution for particular values of h, $n \leqslant 7$

$n = 4$:	$K_{1,1}$	·7	·7879	·8810	1
	$K_{1,2}$	·4	·3030	·1905	0
	$K_{1,3}$	·1	·0303	−·0238	0
	$K_{1,4}$	−·2	−·1212	−·0476	0
	$K_{2,2}$	·3	·4242	·5952	1
	$K_{2,3}$	·2	·2424	·2381	0
$n = 5$:	$K_{1,1}$	·6	·7708	·8803	1
	$K_{1,2}$	·4	·2917	·1880	0
	$K_{1,3}$	·2	·0417	−·0256	0
	$K_{1,4}$	0	−·0417	−·0342	0
	$K_{1,5}$	−·2	−·0625	−·0085	0
	$K_{2,2}$	·3	·4167	·5855	1
	$K_{2,3}$	·2	·25	·2308	0
	$K_{2,4}$	·1	·0833	·0299	0
	$K_{3,3}$	·2	·4167	·5897	1
$n = 6$:	$K_{1,1}$	·5238	·7692	·8803	1
	$K_{1,2}$	·3810	·2885	·1882	0
	$K_{1,3}$	·2381	·0385	−·0251	0
	$K_{1,4}$	·0952	−·0385	−·0338	0
	$K_{1,5}$	−·0476	−·0385	−·0118	0
	$K_{1,6}$	−·1905	−·0192	·0020	0
	$K_{2,2}$	·2952	·4103	·5843	1
	$K_{2,3}$	·2095	·2436	·2275	0
	$K_{2,4}$	·1238	·0897	·0275	0
	$K_{2,5}$	·0381	·0064	−·0157	0
	$K_{3,3}$	·1810	·4103	·5802	1
	$K_{3,4}$	·1524	·2564	·2237	0
$n = 7$:	$K_{1,1}$	·4643	·7692	·8803	1
	$K_{1,2}$	·3571	·2885	·1883	0
	$K_{1,3}$	·25	·0385	−·0249	0
	$K_{1,4}$	·1429	−·0385	−·0332	0
	$K_{1,5}$	·0357	−·0385	−·0114	0
	$K_{1,6}$	−·0714	−·0192	−·0010	0
	$K_{1,7}$	−·1786	0	·0019	0
	$K_{2,2}$	·2857	·4087	·5843	1
	$K_{2,3}$	·2143	·2404	·2274	0
	$K_{2,4}$	·1429	·0865	·0272	0
	$K_{2,5}$	·0714	·0096	−·0159	0
	$K_{2,6}$	0	−·0144	−·0103	0
	$K_{3,3}$	·1786	·4038	·5791	1
	$K_{3,4}$	·1429	·25	·2205	0
	$K_{3,5}$	·1071	·0962	·0251	0
	$K_{4,4}$	·1429	·4038	·5710	1

Table 2 shows that with increasing h, the coefficients in the main diagonal tend to increase towards 1, and the differences between these elements tend to diminish. This equalizing tendency operates quickly with regard to all $K_{i,i}$ except $K_{1,1}$ and also $K_{n,n}$, but very slowly between $K_{1,1}$ and $K_{n,n}$ on the one hand and $K_{i,i}$ on the other; the difference between, say, $K_{1,1}$ and $K_{2,2}$ remains substantial even for $h = 4$. The elements outside the main diagonal gradually tend towards zero, though some of them increase first before diminishing, and others change sign in the process.

3. A Particular Solution

To simplify matters, the analysis will now be confined to the case $h = 1$. This gives us one particular trend, which may be denoted as the central quasi-linear trend. It is obvious from Table 1 that the solution based on $h = 1$ will be less unwieldy than those based on other values of h. Equations (4) reduce to

$$\begin{aligned}
n = 3:\quad 2T_1 - 2T_2 + T_3 &= Y_1,\\
-2T_1 + 5T_2 - 2T_3 &= Y_2,\\
T_1 - 2T_2 + 2T_3 &= Y_3;\\
n \geqslant 4:\quad 2T_1 - 2T_2 + T_3 &= Y_1,\\
-2T_1 + 6T_2 - 4T_3 + T_4 &= Y_2,\\
T_1 - 4T_2 + 7T_3 - 4T_4 + T_5 &= Y_3,\\
\cdots\cdots\cdots\cdots\cdots\cdots&\cdots\cdots\\
T_{n-4} - 4T_{n-3} + 7T_{n-2} - 4T_{n-1} + T_n &= Y_{n-2},\\
T_{n-3} - 4T_{n-2} + 6T_{n-1} - 2T_n &= Y_{n-1},\\
T_{n-2} - 2T_{n-1} + 2T_n &= Y_n. \qquad (4a)
\end{aligned}$$

For $n \leqslant 7$, the solution can immediately be derived from Table 1 by substituting $h = 1$. It is seen that in some cases, D and $DK_{1,1}, DK_{1,2}, \ldots$ will then contain a common factor λ which is a positive integer, and we can replace D by $D' = D/\lambda$. We have the following values:

n	3	4	5	6	7
D	7	33	144	624	2,704
λ	1	1	3	4	13
D'	7	33	48	156	208

The solution can then be written as follows:

$$\begin{aligned}
n = 3:\quad T_1 &= \tfrac{1}{7}(6Y_1 + 2Y_2 - Y_3)\\
T_2 &= \tfrac{1}{7}(2Y_1 + 3Y_2 + 2Y_3);
\end{aligned}$$

$$n = 4: \quad T_1 = \tfrac{1}{33}(26Y_1 + 10Y_2 + Y_3 - 4Y_4)$$
$$T_2 = \tfrac{1}{33}(10Y_1 + 14Y_2 + 8Y_3 + Y_4);$$

$$n = 5: \quad T_1 = \tfrac{1}{48}(37Y_1 + 14Y_2 + 2Y_3 - 2Y_4 - 3Y_5)$$
$$T_2 = \tfrac{1}{48}(14Y_1 + 20Y_2 + 12Y_3 + 4Y_4 - 2Y_5)$$
$$T_3 = \tfrac{1}{48}(2Y_1 + 12Y_2 + 20Y_3 + 12Y_4 + 2Y_5);$$

$$n = 6: \quad T_1 = \tfrac{1}{156}(120Y_1 + 45Y_2 + 6Y_3 - 6Y_4 - 6Y_5 - 3Y_6)$$
$$T_2 = \tfrac{1}{156}(45Y_1 + 64Y_2 + 38Y_3 + 14Y_4 + Y_5 - 6Y_6)$$
$$T_3 = \tfrac{1}{156}(6Y_1 + 38Y_2 + 64Y_3 + 40Y_4 + 14Y_5 - 6Y_6);$$

$$n = 7: \quad T_1 = \tfrac{1}{208}(160Y_1 + 60Y_2 + 8Y_3 - 8Y_4 - 8Y_5 - 4Y_6)$$
$$T_2 = \tfrac{1}{208}(60Y_1 + 85Y_2 + 50Y_3 + 18Y_4 + 2Y_5 - 3Y_6 - 4Y_7)$$
$$T_3 = \tfrac{1}{208}(8Y_1 + 50Y_2 + 84Y_3 + 52Y_4 + 20Y_5 + 2Y_6 - 8Y_7)$$
$$T_4 = \tfrac{1}{208}(-8Y_1 + 18Y_2 + 52Y_3 + 84Y_4 + 52Y_5 + 18Y_6 - 8Y_7). \tag{9}$$

The remaining expressions can be written down immediately on account of symmetry considerations. The exact solution in these cases readily lends itself to practical application.

It is not difficult to solve (4a) for larger values of n. The procedure is simplified by the fact that the last $n-2$ equations in the case of $n+1$ observations are the same as the last $n-2$ equations in the case of n observations, with all suffixes increased by 1. This means that if the solution is obtained for successive values of n, only five linear equations with five unknowns have to be solved in each case.

The solution has been worked out for all $n \leqslant 15$. For $n = 15$, the five equations to be solved are:

$$\begin{aligned}
2T_1 - 2T_2 + T_3 &= Y_1,\\
-2T_1 + 6T_2 - 4T_3 + T_4 &= Y_2,\\
T_1 - 4T_2 + 7T_3 - 4T_4 + T_5 &= Y_3,\\
1{,}519T_2 - 5{,}398T_3 + 8{,}079T_4 - 2{,}000T_5 &= 1{,}519Y_4 + 678Y_5 + 158Y_6 - 38Y_7\\
&\quad - 65Y_8 - 40Y_9 - 15Y_{10} - 2Y_{11}\\
&\quad + 2Y_{12} + 2Y_{13} + Y_{14},\\
-1{,}038T_2 + 4{,}315T_3 - 7{,}556T_4 + 4{,}079T_5 &= -1{,}038Y_4 + 163Y_5 + 362Y_6 + 234Y_7\\
&\quad + 92Y_8 + 15Y_9 - 10Y_{10} - 11Y_{11}\\
&\quad - 6Y_{12} - 2Y_{13} + Y_{15}.
\end{aligned}$$

Multiplying the third and fourth equations by 2,000 and 1 respectively and adding, then the third and fifth by $-4{,}079$ and 1 respectively and adding, we obtain

$$2{,}000T_1 - 6{,}481T_2 + 8{,}602T_3 + 79T_4 = 2{,}000Y_3 + 1{,}519Y_4 + \ldots + 2Y_{13} + Y_{14},$$

$$-4{,}079T_1 + 15{,}278T_2 - 24{,}238T_3 + 8{,}760T_4 = -4{,}079Y_3 - 1{,}038Y_4 + \ldots - 2Y_{13} + Y_{15}.$$

These equations, together with the first and second, lead to the solution for $n = 15$. In addition, they could be utilized when $n = 16$, which would thus call for the solution of the five equations

$$\begin{aligned}
2T_1 - 2T_2 + T_3 &= Y_1,\\
-2T_1 + 6T_2 - 4T_3 + T_4 &= Y_2,\\
T_1 - 4T_2 + 7T_3 - 4T_4 + T_5 &= Y_3,\\
2{,}000T_2 - 6{,}481T_3 + 8{,}602T_4 + 79T_5 &= 2{,}000Y_4 + 1{,}519Y_5 + 678Y_6 + 158Y_7\\
&\quad - 38Y_8 - 65Y_9 - 40Y_{10} - 15Y_{11}\\
&\quad - 2Y_{12} + 2Y_{13} + 2Y_{14} + Y_{15},\\
-4{,}079T_2 + 15{,}278T_3 - 24{,}238T_4 + 8{,}760T_5 &= -4{,}079Y_4 - 1{,}038Y_5 + 163Y_6 + 362Y_7\\
&\quad + 234Y_8 + 92Y_9 + 15Y_{10} - 10Y_{11}\\
&\quad - 11Y_{12} - 6Y_{13} - 2Y_{14} + Y_{16}.
\end{aligned}$$

The procedure incidentally yields the exact value of

$$D = \begin{vmatrix}
2 & -2 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0\\
-2 & 6 & -4 & 1 & 0 & \ldots & 0 & 0 & 0 & 0 & 0\\
1 & -4 & 7 & -4 & 1 & \ldots & 0 & 0 & 0 & 0 & 0\\
 & & \ldots & \ldots & \ldots & & \ldots & \ldots & \ldots & & \\
0 & 0 & 0 & 0 & 0 & \ldots & 1 & -4 & 7 & -4 & 1\\
0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & -4 & 6 & -2\\
0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 & -2 & 2
\end{vmatrix}.$$

Again, D and $DK_{1,1}, DK_{1,2}, \ldots$ in some cases contain a common factor λ, and in writing down the solution, D can be replaced by D'.

The following tabulation gives, for $8 \leqslant n \leqslant 15$, the values of D, λ, D' and those values of $D' K_{i,j}$ which are defined, excluding those that can be derived from symmetry considerations.

TABLE 3

Solution of equations (4*a*) *for* $8 \leqslant n \leqslant 15$

n	8	9	10	11	12	13	14	15
D	11,713	50,727	219,681	951,360	4,120,000	17,842,240	77,268,321	334,621,287
λ	1	1	3	1	40	1	13	9
D'	11,713	50,727	73,227	951,360	103,000	17,842,240	5,943,717	37,180,143
$D' K_{1,1}$	9,009	39,014	56,318	731,679	79,216	13,722,240	4,571,237	28,594,774
$D' K_{1,2}$	3,380	14,638	21,130	274,518	29,721	5,148,440	1,715,080	10,728,458
$D' K_{1,3}$	455	1,975	2,851	37,038	4,010	694,640	231,403	1,447,511
$D' K_{1,4}$	−442	−1,900	−2,740	−35,598	−3,854	−667,600	−222,394	−1,391,156
$D' K_{1,5}$	−442	−1,887	−2,715	−35,265	−3,818	−661,360	−220,314	−1,378,143
$D' K_{1,6}$	−234	−986	−1,406	−18,240	−1,975	−342,120	−113,966	−712,894
$D' K_{1,7}$	−65	−310	−430	−5,535	−600	−104,000	−34,645	−216,710
$D' K_{1,8}$	52	14	2	78	7	1,000	320	2,002
$D' K_{1,9}$		169	109	1,362	146	24,880	8,245	51,545
$D' K_{1,10}$			108	1,002	110	18,640	6,126	38,220
$D' K_{1,11}$				321	46	8,400	2,714	16,783
$D' K_{1,12}$					−9	1,640	634	3,770
$D' K_{1,13}$						−2,560	−203	−1,066
$D' K_{1,14}$							−520	−1,742
$D' K_{1,15}$								−1,209
$D' K_{2,2}$	4,785	20,723	29,914	388,636	42,076	7,288,641	2,428,037	15,188,267
$D' K_{2,3}$	2,810	12,170	17,568	228,236	24,710	4,280,402	1,425,914	8,919,618
$D' K_{2,4}$	1,003	4,345	6,275	81,524	8,826	1,588,882	509,311	3,185,929
$D' K_{2,5}$	102	444	650	8,470	917	158,838	52,914	331,000
$D' K_{2,6}$	−158	−682	−967	−12,480	−1,350	−233,855	−77,902	−487,298
$D' K_{2,7}$	−144	−626	−886	−11,350	−1,225	−212,160	−70,676	−442,094
$D' K_{2,8}$		−299	−450	−5,684	−608	−105,185	−35,045	−219,219
$D' K_{2,9}$			−116	−1,676	−174	−29,878	−9,970	−62,410
$D' K_{2,10}$				164	10	1,998	633	3,815
$D' K_{2,11}$					51	8,558	2,818	17,348
$D' K_{2,12}$						5,919	2,006	12,266
$D\ K_{2,13}$							800	5,298
$D' K_{2,14}$								907
$D' K_{3,3}$	4,710	20,390	29,434	382,396	41,400	7,171,524	2,389,022	14,944,214
$D' K_{3,4}$	2,890	12,490	18,030	234,244	25,360	4,392,964	1,463,410	9,154,170
$D' K_{3,5}$	1,088	4,662	6,730	87,470	9,470	1,640,396	546,456	3,418,286
$D' K_{3,6}$	152	608	878	11,520	1,250	216,530	72,128	451,192
$D' K_{3,7}$		−632	−912	−11,630	−1,250	−216,320	−72,062	−450,768
$D' K_{3,8}$			−904	−11,524	−1,230	−212,370	−70,730	−442,442
$D' K_{3,9}$				−6,076	−640	−109,516	−36,430	−227,910
$D' K_{3,10}$					−200	−33,284	−10,986	−68,810
$D' K_{3,11}$						316	208	1,130
$D' K_{3,12}$							2,744	16,992
$D' K_{3,13}$								12,728
$D' K_{4,4}$	4,658	20,090	28,990	376,636	40,776	7,063,364	2,352,986	14,718,794
$D' K_{4,5}$	2,856	12,210	17,590	228,530	24,742	4,285,836	1,427,712	8,930,858
$D' K_{4,6}$		4,552	6,502	84,480	9,150	1,585,010	527,992	3,302,768
$D' K_{4,7}$			808	10,510	1,150	199,680	66,518	416,072
$D' K_{4,8}$				−11,836	−1,258	−216,370	−72,010	−450,450
$D' K_{4,9}$					−1,224	−209,036	−69,410	−434,090
$D' K_{4,10}$						−107,844	−35,490	−221,690
$D' K_{4,11}$							−10,648	−66,002
$D' K_{4,12}$								1,912

TABLE 3 (*continued*)

Solution of equations (4*a*) *for* $8 \leqslant n \leqslant 15$

n	8	9	10	11	12	13	14	15
D	11,713	50,727	219,681	951,360	4,120,000	17,842,240	77,268,321	334,621,287
λ	1	1	3	1	40	1	13	9
D'	11,713	50,727	73,227	951,360	103,000	17,842,240	5,943,717	37,180,143
$D' K_{5,5}$		19,869	28,565	370,975	40,164	6,957,284	2,317,626	14,497,573
$D' K_{5,6}$			17,400	225,600	24,425	4,231,030	1,409,430	8,816,430
$D' K_{5,7}$				83,585	9,050	1,568,320	522,447	3,267,998
$D' K_{5,8}$					1,139	199,370	66,570	416,416
$D' K_{5,9}$						−213,924	−70,755	−442,175
$D' K_{5,10}$							−68,652	−428,050
$D' K_{5,11}$								−219,309
$D' K_{6,6}$				369,600	40,000	6,929,025	2,308,190	14,438,410
$D' K_{6,7}$					24,375	4,222,400	1,406,590	8,798,510
$D' K_{6,8}$						1,567,775	522,475	3,268,265
$D' K_{6,9}$							67,100	420,700
$D' K_{6,10}$								−439,425
$D' K_{7,7}$						6,926,400	2,307,365	14,433,110
$D' K_{7,8}$							1,406,600	8,798,790
$D' K_{7,9}$								3,269,575
$D' K_{8,8}$								14,433,419

The exact solution would be cumbersome to use in practical applications for $n \geqslant 8$. It is, however, possible to give an approximate solution which remains valid irrespective of n from a minimum value onwards. It can be written as follows:

$$\left.\begin{aligned}
T_1 &= \cdot 77Y_1+\cdot 29Y_2+\cdot 04Y_3-\cdot 04Y_4-\cdot 04Y_5-\cdot 02Y_6 && (n\geqslant 8)\\
T_2 &= \cdot 29Y_1+\cdot 41Y_2+\cdot 24Y_3+\cdot 08Y_4+\cdot 01Y_5-\cdot 01Y_6-\cdot 02Y_7 && (n\geqslant 8)\\
T_3 &= \cdot 04Y_1+\cdot 24Y_2+\cdot 40Y_3+\cdot 25Y_4+\cdot 09Y_5+\cdot 01Y_6-\cdot 01Y_7-\cdot 02Y_8 && (n\geqslant 8)\\
T_4 &= \begin{cases} -\cdot 04Y_1+\cdot 08Y_2+\cdot 25Y_3+\cdot 40Y_4+\cdot 25Y_5+\cdot 09Y_6+\cdot 01Y_7-\cdot 04Y_8 & (n=8)\\ -\cdot 04Y_1+\cdot 08Y_2+\cdot 25Y_3+\cdot 40Y_4+\cdot 24Y_5+\cdot 09Y_6+\cdot 01Y_7 & \\ \qquad -\cdot 01Y_8-\cdot 02Y_9 & (n\geqslant 9)\end{cases}\\
T_5 &= \begin{cases} -\cdot 04Y_1+\cdot 01Y_2+\cdot 09Y_3+\cdot 24Y_4+\cdot 40Y_5+\cdot 24Y_6+\cdot 09Y_7 & \\ \qquad +\cdot 01Y_8-\cdot 04Y_9 & (n=9)\\ -\cdot 04Y_1+\cdot 01Y_2+\cdot 09Y_3+\cdot 24Y_4+\cdot 39Y_5+\cdot 24Y_6+\cdot 09Y_7 & \\ \qquad +\cdot 01Y_8-\cdot 01Y_9-\cdot 02Y_{10} & (n\geqslant 10)\end{cases}\\
T_i &= -\cdot 02Y_{i-5}-\cdot 01Y_{i-4}+\cdot 01Y_{i-3}+\cdot 09Y_{i-2}+\cdot 24Y_{i-1}+\cdot 38Y_i\\
&\qquad +\cdot 24Y_{i+1}+\cdot 09Y_{i+2}+\cdot 01Y_{i+3}-\cdot 01Y_{i+4}-\cdot 02Y_{i+5} && (i\geqslant 6,\ n\geqslant i+5)
\end{aligned}\right\} \quad (10)$$

It is easily seen that equations (4a) are approximately satisfied by (10). For example, we obtain from (10) for $n = 8$,

$$2T_1 - 2T_2 + T_3 = Y_1 + \cdot 01Y_4 - \cdot 01Y_5 - \cdot 01Y_6 + \cdot 03Y_7 - \cdot 02Y_8,$$

and the expression on the right-hand side does not differ much from that in (4a), which is Y_1.

The approximate solution (10) possesses the symmetry property (6) and the additivity property (7), but property (8) does not hold exactly.

It should be noted that the formulae (10) cannot be used to derive T_5 for $n < 9$, T_6 for $n < 10$, etc. The expressions for T_{n-3}, T_{n-2}, T_{n-1} and T_n are deduced with the aid of symmetry considerations from the expressions for T_1, T_2, T_3 and T_4; the expression for T_{n-4} also from that for T_5 if $n \geqslant 10$.

For a large number of observations, the central quasi-linear trend is thus approximated by an eleven-year weighted moving average with negative weights for the first two and the last two years used, together with special formulae for the first five and the last five years of the observation period.

The use of approximation (10) offers, of course, great computational advantages, not least the fact that when a further observation is added, only the last five trend values have to be recalculated in addition to the new trend value to be computed. In the absence of other information, the last two trend values obtained may also be used for predicting a continuation of the most recently experienced trend direction, that is to say,

$$T_{n+1(\text{pred})} = 2T_n - T_{n-1}, \quad T_{n+2(\text{pred})} = 3T_n - 2T_{n-1}, \ldots.$$

It would, of course, not be advisable to carry the prediction too far.

4. EXAMPLES

The method developed here will now be illustrated with the aid of two examples A and B, the first one furnished by an artificially constructed series, the other one by actual data. In both cases, the approximation formulae (10) have been used for computing the trend values; owing to this fact, there are slight differences between the values obtained for $g_{i-1} - 2g_i + g_{i+1}$ and e_i, in particular between the values of g_2 and e_1 and between the values of g_{n-1} and e_n. The permanent disturbances g_i are less accurately calculated than the trend values T_i and temporary disturbances e_i.

Example A. We take $n = 15$; $h = 1$,

$$Y_i = \eta_i + \delta_i \quad (i = 1, 2, \ldots, 15),$$

where $$\eta_i = X_i^2 - 10X_i + 35 \quad (X_i = 1, 2, \ldots, 15),$$

and δ_i is an error term drawn at random from a binomial probability distribution with mean 0 and variance 25, that is to say,

x	-10	-5	0	5	10
probability	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$.

Table 4 shows the construction of the time series, the trend arrived at by fitting a second-degree polynomial, and the solution given by the quasi-linear trend. Given Y_i, the second-degree curve is

$$Y_c = 35\cdot 67 + 6\cdot 196\xi_1' + \cdot 3111\xi_2',$$

where ξ_1', ξ_2' are the tabulated orthogonal polynomials, or

$$Y_{ic} = \cdot 9333X_i^2 - 8\cdot 737X_i + 28\cdot 41.$$

Neither Y_{ic} nor T_i reproduces η_i exactly; Y_{ic} is somewhat closer than T_i, since

$$\sum_{i=1}^{15} (Y_{ic} - \eta_i)^2 = 88, \quad \sum_{i=1}^{15} (T_i - \eta_i)^2 = 135,$$

and incidentally, $$\sum_{i=1}^{15} (Y_{ic} - \eta_i) = \sum_{i=1}^{15} (T_i - \eta_i) = -30.$$

TABLE 4

Artificially constructed time series

i	η_i	δ_i	Y_i	Y_{ic}	T_i	e_i	g_i
1	26	− 5	21	20·6	18·3	+2·7	
2	19	−10	9	14·7	14·2	−5·2	+3·0
3	14	0	14	10·6	13·1	+ ·9	− ·2
4	11	+ 5	16	8·4	11·8	+4·2	−1·7
5	10	0	10	8·1	8·8	+1·2	+1·7
6	11	−10	1	9·6	7·5	−6·5	+4·6
7	14	0	14	13·0	10·8	+3·2	+2·5
8	19	− 5	14	18·2	16·6	−2·6	+3·0
9	26	0	26	25·4	25·4	+ ·6	+1·3
10	35	+ 5	40	34·4	35·5	+4·5	− ·2
11	46	− 5	41	45·2	45·4	−4·4	+3·2
12	59	0	59	58·0	58·5	+ ·5	+2·3
13	74	0	74	72·6	73·9	+ ·1	+ ·6
14	91	0	91	89·0	89·7	+1·3	− ·1
15	110	− 5	105	107·4	105·4	− ·4	

The largest discrepancies between η_i and T_i are at both ends of the period. The quantities T_1 and T_{15} are heavily influenced by Y_1 and Y_{15} and thus, as it were, not able to pick out the "true" trend.

On the other hand, T_i gives a closer fit to Y_i than Y_{ic}, since

$$\sum_{i=1}^{15} (Y_i - T_i)^2 = 156, \quad \sum_{i=1}^{15} (Y_i - Y_{ic})^2 = 260.$$

As $$\sum_{i=1}^{15} (Y_i - \bar{Y})^2 = 14{,}605,$$

1·0 per cent. of the variation in Y remains unexplained by the quasi-linear trend, as against 1·8 per cent. unexplained by the parabolic regression curve.

Example B. Here $n = 10$, $h = 1$, and Y_i represents the income of farms in Australia (£m.) for each financial year from 1948/49 to 1957/58, taken from Table I, Appendix C, p. 16, of *National Income and Expenditure*, 1958–59. The series has been chosen because of its large fluctuations.

In addition to T_i, a fourth-degree polynomial denoted by Y_{ic} has been fitted to the data. The addition of a fifth-degree term would improve the fit only slightly;

the third- and fourth-degree terms improve the fit substantially, even though, on account of a large residual variance, the statistical significance cannot be established.

Both methods yield practically the same residual sum of squares

$$\sum_{i=1}^{10} (Y_i - Y_{ic})^2 = 58{,}923, \quad \sum_{i=1}^{10} (Y_i - T_i)^2 = 59{,}069,$$

and since

$$\sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 137{,}795,$$

the unexplained portion of the variance is about 43 per cent. whichever trend is used.

TABLE 5

Income of farms, Australia, 1948/49—1957/58

i	Y_i	Y_{ic}	T_i	e_i	g_i
1	321	297	357	− 36	
2	448	544	485	− 37	− 37
3	756	603	576	+180	−107
4	441	571	560	−119	− 3
5	572	516	541	+ 31	− 15
6	499	479	507	− 8	+ 4
7	447	472	477	− 30	+ 11
8	443	481	458	− 15	+ 4
9	519	465	442	+ 77	− 49
10	335	353	377	− 42	

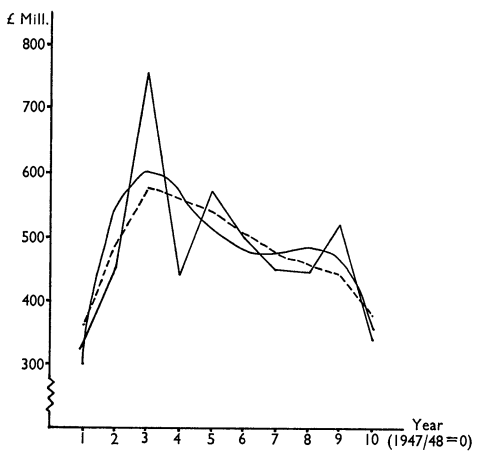

FIG. 1. Income of farms, Australia.

In Fig. 1 the data are connected by solid lines and the quasi-linear trend values by broken lines; the fourth-degree curve is also shown. It is seen that for both the initial and the final year, 1948/49 and 1957/58, in both of which farm income was low, the curve falls below the quasi-linear trend, and for 1948/49 it even falls below the low actual figure. The quasi-linear trend, quite apart from scoring heavily with regard to computational ease, thus seems the more suitable one of the two trends, as it is more successful in smoothing out the fluctuations in the data than the regression curve.

The White Paper gives £408m. as a provisional estimate for the farm income of 1958/59. If it is desired to add $Y_{11} = 408$ to the observations, the terms from T_6 onward are affected, but only T_{10} substantially. We now obtain

$$T_6 = 504, \quad T_7 = 474, \quad T_8 = 456,$$

$$T_9 = 446, \quad T_{10} = 404, \quad T_{11} = 386.$$

The trend still maintains a gradual fall after the sharp rise up to 1950/51, but the fall is mitigated for recent years, compared with the trend based on ten observations.

REFERENCES

HENDERSON, R. (1924), "A new method of graduation", *Trans. Actuar. Soc. Amer.*, **25**, 29–40.
KENDALL, M. G. (1948), *The Advanced Theory of Statistics*, Vol. 2, 2nd ed. London: Griffin.
MACAULAY, F. R. (1931), *The Smoothing of Time Series*. New York: National Bureau of Economic Research.
WHITTAKER, E. T. (1923), "On a new method of graduation", *Proc. Edinb. math. Soc.*, **41**, 63–75.
—— (1924), "On the theory of graduation", *Proc. R. Soc. Edinb.*, **44**, 77–83.
—— and ROBINSON, G. (1926), *The Calculus of Observations*, 2nd ed. London: Blackie.