Steffen L. Lauritzen

Time Series Analysis in 1880. A Discussion of Contributions made by T. N. Thiele

Preprint
March
1981
3

Institute of Mathematical Statistics
University of Copenhagen

Steffen L. Lauritzen

TIME SERIES ANALYSIS IN 1880.
A DISCUSSION OF CONTRIBUTIONS MADE
BY T.N. THIELE.

Preprint 1981 No. 3

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

March 1981

Abstract. We describe and discuss a paper of T.N. Thiele from 1880 where he formulates and analyses a model for a time series consisting of a sum of a regression component, a Brownian motion and a white noise. He derives a recursive procedure for estimating the regression component and predicting the Brownian motion. The procedure is now known as Kalman filtering. He estimates the unknown variances of the Brownian motion and the white noise by an iterative procedure that essentially is the EM-algorithm. We finally give a short account of an application of Thiele's model and method to the description of hormone production during normal pregnancy.

Key words: ARIMA, Brownian motion, EM-algorithm, incomplete data, Kalman filtering, restricted maximum likelihood, time series, variance components.

1. Introduction

In the present paper we shall give a description and a discussion of a paper by T.N. Thiele (1880, 1880a) on a particular time series model used by him in a problem of astronomical geodesy, more precisely in connection with the problem of determining the distance from Copenhagen to Lund (Sweden).

Although the paper has been overlooked by today's statisticians, it contains remarkable results, results that are interesting even today and not just from a historical point of view.

A short discussion of Thiele's model and method has survived in the sense that it is described in the textbook by Helmert (1907) that has been used as a basis for teaching statistics to geodesists until recent time.

Thiele proposes a model consisting of a sum of a regression component, a Brownian motion and a white noise for his observations, although he does not use these terms himself.

He solves the problem of estimating the regression coefficients and predicting the values of the Brownian motion by the method of least squares and gives an elegant recursive procedure for carrying out the calculations. The procedure is nowadays known as Kalman filtering (Kalman and Bucy, 1961).

The iterative procedure used by Thiele to estimate the variances of the Brownian motion and the noise is related to the EM-algorithm described by Dempster, Laird and Rubin (1977) or, more precisely, identical to the algorithm given by Patterson and Thompson (1971) for variance component models.

Thiele did not derive the distribution of his variance estimates which is rather typical for statistical work at that time. In later work by Thiele he becomes interested in such problems but in this particular paper they seem beyond his horizon.

It is perhaps even more remarkable that this paper is the first paper written by Thiele on the method of least squares!

There are obvious reasons for his paper to have been more or less neglected by other statisticians. Thiele is certainly not friendly to his readers and assumes these to have quite an exceptional knowledge and understanding of Gauss' method of least squares. His ideas seem to be so much ahead of his time (100 years) that his contemporaries did not have a chance to understand the paper and, maybe more important, to grasp the significance of the work. When later the time was ripe, the development of statistics was so much concentrated in England and the U.S. of A., where no one seemingly would dream of looking for essential contributions to statistics made by a Danish astronomer in 1880.

My interest in this work arose partly from reading a version of the paper by Hald (1981) containing a short description of Thiele's paper, and partly because I for some time had been working with statistical description of hormone concentrations in plasma during pregnancy, where the data seemed to be described perfectly by Thiele's model. In section 7 we shall give a short description of the experiences in applying Thiele's procedure to that problem.

The "quotations" from Thiele's paper given here are not direct translations from the original paper, but made such as to convey the meaning, atmosphere and writing style of Thiele although modern notation and concepts are used.

2. The model

Before we proceed to discuss the statistical analysis performed by Thiele, we shall briefly sketch how he arrives at his model.

Thiele wants to give a model describing the observation errors from a sequence of measurements obtained through time.

He discusses first the empirical fact that such errors often appear as if they had a systematic component but emphasizes that this is not true since no procedure of correction seems to remove the phenomenon. Thus another explanation must be appropriate and he attributes the phenomenon to the fact that a (random) component of the errors is accumulated through time.

More precisely he considers measurements made by an instrument where part of the error is due to fluctuations of the position of the instrument itself. If $X(t)$ is the position of the instrument at time t, the most likely position of the instrument at time $t+\Delta t$ should be the position immediately before, i.e. $X(t)$, and deviations from this should be governed by the normal distribution law. He then concludes that any sequence of instrument positions $X(t_0),\ldots,X(t_n)$, where $t_0,\ldots,t_n$ are consecutive time points, should have the property that the increments are independent, normally distributed with

$$E(X(t_i)-X(t_{i-1}))=0$$

$$V(X(t_i)-X(t_{i-1}))=\int_{t_{i-1}}^{t_i}\omega^2(u)\,du=\omega_i^2\ ,$$

where $\omega^2(u)$ is a function describing the average size of the square of the fluctuations at time u.

In the special case where $\omega^2(u) \equiv \omega^2$ we note that $X(t)$ is what today is known as a Wiener process or Brownian motion.

The 'quasi-systematic' variation in the errors is then supposed to be due to a process of the above type.

The observations themselves are now supposed to be independent, normally distributed around the X-values. More precisely, for $i = 0, 1, \ldots, n$ let

$$Z(t_i) = X(t_i) + \varepsilon(t_i)$$

where $\varepsilon(t_0), \varepsilon(t_1), \ldots, \varepsilon(t_n)$ are independent and independent of the X-process, normally distributed with expectation equal to zero and

$$V(\varepsilon(t_i)) = \sigma_i^2 \;.$$

I have deliberately not specified the joint distribution of all the variables completely ($X(t_0)$ - s distribution is unspecified). This is because Thiele does not either. We shall later see that, in fact, $X(t_0)$ plays the role of what we today would call a parameter, in Thiele's statistical analysis although Thiele does not make a clear distinction between a parameter and an unobserved random variable.

Note the special case of the model obtained by assuming

a) equidistant time points: $t_i = i$

b) constant variance in the time fluctuations: $\omega^2(u) = \omega^2$

c) constant measurement error: $\sigma_i^2 = \sigma^2$.

We then get for the differenced process

$$\nabla Z(i) = Z(t_i) - Z(t_{i-1})$$

that this is a stationary process with expectation equal to zero and covariance function

$$r(k) = V(\nabla Z(i), \nabla Z(i+k))$$

$$= \begin{cases} \omega^2 + \sigma^2 & \text{if } k = 0 \\ -\sigma^2 & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

which is a moving average process of order one. In other words, Z(0),Z(1),...,Z(n) is a sample from an ARIMA(0,1,1) process.

An ARIMA (0,1,1) process <u>even with missing observations</u> is therefore a special case of Thiele's model and can be treated with Thiele's methods, to be described subsequently.

3. Least squares prediction of the Brownian motion

In Thiele's formulation the primary objective is to estimate the unknown values of $X(t_0),\dots,X(t_n)$. As earlier mentioned he does not distinguish between a parameter and an unobserved random variable but treats these unknown quantities seemingly alike. Let us examine his procedure in some detail.

First, at this stage, Thiele considers σ_i^2 and ω_i^2 known, $X(t_0),\dots,X(t_n)$ unknown and $Z(t_0),\dots,Z(t_n)$ known, i.e. observed.

He then writes:

"We get the following system of $2n+1$ equations with $n+1$ unknowns:

$$\left.\begin{array}{lll}
X(t_1) - X(t_0) & = 0 & \text{with weight } \omega_1^{-2} \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
X(t_n) - X(t_{n-1}) & = 0 & \text{with weight } \omega_n^{-2} \\
Z(t_0) - X(t_0) & = 0 & \text{with weight } \sigma_0^{-2} \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
Z(t_n) - X(t_n) & = 0 & \text{with weight } \sigma_n^{-2}
\end{array}\right\} \quad (3.1)$$

Solving these by the method of least squares leads to the system of $n+1$ equations:

$$\left.\begin{aligned}
\sigma_0^{-2}\, Z(t_0) &= (\sigma_0^{-2} + \omega_1^{-2})\hat{X}(t_0) - \omega_1^{-2}\hat{X}(t_1) \\
\sigma_1^{-2}\, Z(t_1) &= -\omega_1^{-2}\hat{X}(t_0) + (\omega_1^{-2} + \sigma_1^{-2} + \omega_2^{-2})\,\hat{X}(t_1) - \omega_2^{-2}\hat{X}(t_2) \\
&\;\;\vdots \qquad\qquad\qquad\qquad \vdots \\
\sigma_i^{-2}\, Z(t_i) &= -\omega_i^{-2}\hat{X}(t_{i-1}) + (\omega_i^{-2} + \sigma_i^{-2} + \omega_{i+1}^{-2})\,\hat{X}(t_i) - \omega_i^{-2}\hat{X}(t_{i+1}) \\
&\;\;\vdots \qquad\qquad\qquad\qquad \vdots \\
\sigma_n^{-2}\, Z(t_n) &= -\omega_n^{-2}\hat{X}(t_{n-1}) + (\omega_n^{-2} + \sigma_n^{-2})\,\hat{X}(t_n)
\end{aligned}\right\} \quad (3.2)$$

which we shall now show how to solve" .

Thiele's argument <u>is</u> as short as this, showing how he (of course) assumed the reader to be absolutely familiar with the method of least squares.

Before we proceed to describe Thiele's recursive procedure we shall discuss in which sense the estimates $\hat{X}(t_i)$ given by (3.2) (or <u>predictions</u>, as we would say today) are the 'right' ones.

A rapid check will show that the values $\hat{X}(t_0),\ldots,\hat{X}(t_n)$ given by (3.2) minimize the quadratic form

$$Q = \sum_{i=0}^{n} \sigma_i^{-2}(Z(t_i) - X(t_i))^2 + \sum_{i=1}^{n} \omega_i^{-2}\,(X(t_i) - X(t_{i-1}))^2 \quad .$$

Apart from an additive constant we have

$$Q = -\ 2 \log f \ ,$$

where f is the joint density of $X(t_1),\ldots,X(t_n),Z(t_0),\ldots,Z(t_n)$ where $X(t_0)$ is considered non-random, i.e. a <u>parameter</u>.

For the sake of clarity we shall henceforth write $X(t_0) = \alpha$ and realise that the joint distribution of $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ where

$$\underset{\sim}{X} = (X(t_1), \ldots, X(t_n))'$$

$$\underset{\sim}{Z} = (Z(t_0), \ldots, Z(t_n))'$$

is multivariate normal with

$$E\underset{\sim}{X} = (\alpha, \ldots, \alpha)' \quad E\underset{\sim}{Z} = (\alpha, \ldots, \alpha)'$$

and the joint density of $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ can therefore be factorized as

$$f(\underset{\sim}{X}, \underset{\sim}{Z}; \alpha) = h(\underset{\sim}{Z}; \alpha)\, g(\underset{\sim}{X} | \underset{\sim}{Z}; \alpha)$$

into the product of the marginal density of $\underset{\sim}{Z}$ and the conditional density of $\underset{\sim}{X}$ given $\underset{\sim}{Z}$.

For each fixed value of α, the second factor is maximized as a function of $\underset{\sim}{X}$ when

$$\underset{\sim}{X}^* = E_\alpha(\underset{\sim}{X} | \underset{\sim}{Z}) \quad .$$

The maximal value of g will be equal to

$$(2\pi)^{-\frac{n+1}{2}} |\Sigma|^{-\frac{1}{2}}$$

where Σ is the conditional covariance matrix of $\underset{\sim}{X}$ for given $\underset{\sim}{Z}$. Since this does not depend on α, the <u>maximal value</u> does not depend on α and f can thus be maximized as a function of α <u>and</u> $\underset{\sim}{X}$ by letting $\hat{\alpha}$ maximize $h(Z;\alpha)$ and letting

$$\hat{\underset{\sim}{X}} = E_{\hat{\alpha}}(\underset{\sim}{X} | \underset{\sim}{Z}) \quad ,$$

i.e. for $i = 1, \ldots, n$ we have for the solutions to (3.2)

$$\hat{X}(t_i) = E_{\hat{\alpha}}(X(t_i) \mid Z(t_0),\ldots,Z(t_n)) \tag{3.3}$$

and

$$\hat{X}(t_0) = \hat{\alpha}$$

where $\hat{\alpha}$ is the maximum likelihood estimate of α based on $Z(t_0),\ldots,Z(t_n)$.

Another argument, see e.g. Rao (1973), 4a.11, shows that the 'estimator' of $X(t_0),\ldots,X(t_n)$ obtained by minimizing Q is least squares in the sense that for all linear combinations with coefficients $\lambda_0,\ldots,\lambda_n$

$$E\Big(\sum_{j=0}^{n} \lambda_j X(t_j) - \sum_{j=0}^{n} \lambda_j \hat{X}(t_j)\Big)^2$$
$$\geq E\Big(\sum_{j=0}^{n} \lambda_j X(t_j) - k(Z(t_0),\ldots,Z(t_n))\Big)^2$$

for all measurable functions k.

It is worth noting that Thiele also considers the problem of 'estimating' the value of X(s) at a time s, where <u>no</u> observation has been made. He shows correctly that this is given as

$$\hat{X}(s) = \begin{cases} \hat{X}(t_0) & \text{if } s \leqq t_0 \\ \dfrac{(t_{i+1} - s)\hat{X}(t_{i+1}) + (s - t_i)\hat{X}(t_i)}{t_{i+1} - t_i} & \text{if } t_i \leqq s \leqq t_{i+1} \\ \hat{X}(t_n) & \text{if } s \geqq t_n \end{cases} \tag{3.5}$$

His method of obtaining this result is ingenious and elegant and also typical for his work. We shall therefore describe his argument:

The situation where we have <u>not</u> observed X(s) must be equivalent to the one where we introduce a fictitious observation and claim that we have an observation of X(s) with infinite variance. That is we define

$$Z(s) = z ,$$

where z is arbitrary but finite and assume that

$$\sigma^2 = E(Z(s) - X(s))^2 = \infty .$$

Calculate now $\hat{X}(s) = "E(X(s) \mid Z(t_0),\ldots,Z(t_n),Z(s))"$ using the usual system of equations (3.2) where of course $\sigma^{-2} = 0$. This then leads to (3.5).

This "method of the fictitious observation" appears in different versions in other parts of his work as an elegant and useful trick, cf. Hald (1981).

4. Recursive solution to the prediction problem

With the computer capacity of 1880 it was of extreme importance to find a computationally simple way of solving the equations (3.2).

In the numerical example treated by Thiele 74 observations are recorded and without a procedure utilising the relative simple structure in the equations, the computational work involved would be prohibitive.

Thiele solves the problem by giving an elegant recursive procedure for the computations. The procedure consists of two parts:

Part I. Define the following set of coefficients:

$$\left.\begin{aligned} & u_0 = \sigma_0^{-2} \\ & \text{and for } i = 1,\ldots,n \text{ let} \\ & w_i = (\omega_i^2 + u_{i-1}^{-1})^{-1} \\ & u_i = \sigma_i^{-2} + w_i \end{aligned}\right\} \quad . \qquad (4.1)$$

If we now let $X^*(t_i)$ be the best predictor of $X(t_i)$ when only $Z(t_0),\ldots,Z(t_i)$ have been observed, i.e.

$$X^*(t_i) = E_{\hat{\alpha}(i)}(X(t_i) \mid Z(t_0),\ldots,Z(t_i))$$

where $\hat{\alpha}(i)$ is the maximum likelihood estimate of α based on $Z(t_0),\ldots,Z(t_i)$, we have the following recursion formula:

$$\left.\begin{aligned} X^*(t_0) &= Z(t_0) \\ u_{i+1}X^*(t_{i+1}) &= w_{i+1}X^*(t_i) + \sigma_{i+1}^{-2}Z(t_{i+1}) \quad . \end{aligned}\right\} \qquad (4.2)$$

That this is correct is shown by Thiele by an induction argument demonstrating that $X^*(t_{i+1})$ fits into the equations (3.2) if $X^*(t_j)$ $j \leq i$ do.

The computations are indeed very simple to carry out using just a table of reciprocals and a calculator with multiplication and addition.

As (4.2) says, the new predictor is a weighted average of the old predictor and the new observation.

Thiele does not directly give an intuitive 'justification' of formula (4.2) but from other parts of his paper it seems clear that his argument basically must be as follows.

At each stage $X^*(t_i)$ is the best possible measurement of the quantity $X(t_i)$ that can be obtained from $Z(t_0),\ldots,Z(t_i)$. Since the increments of the X-process have expectation equal to zero, $X^*(t_i)$ is also a best measurement of $X(t_{i+1})$ based on $Z(t_0),\ldots,Z(t_i)$. $Z(t_{i+1})$ is also a measurement of $X(t_{i+1})$ and the corresponding observation errors are independent. Thus the best way of combining these is to calculate the weighted average of $X^*(t_i)$ and $Z(t_{i+1})$ with the reciprocal variances as weights. Let now

$$\psi_i = V(X^*(t_i) - X(t_i))$$

$$\phi_{i+1} = V(X^*(t_i) - X(t_{i+1}))$$

Then

$$\phi_{i+1} = V(X^*(t_i) - X(t_i) + X(t_i) - X(t_{i+1})) = \psi_i + \omega_{i+1}^2$$

and thus

$$X^*(t_{i+1}) = \frac{\phi_{i+1}^{-1} X^*(t_i) + \sigma_{i+1}^{-2} Z(t_{i+1})}{\phi_{i+1}^{-1} + \sigma_{i+1}^{-2}} . \tag{4.3}$$

Further we get for the variance of the error of this average that

$$\psi_{i+1} = V(X^*(t_{i+1}) - X(t_{i+1}))$$

$$= \frac{\phi_{i+1}^{-2}\phi_{i+1} + \sigma_{i+1}^{-4}\sigma_{i+1}^{2}}{(\phi_{i+1}^{-1} + \sigma_{i+1}^{-2})}$$

$$= (\phi_{i+1}^{-1} + \sigma_{i+1}^{-2})^{-1} .$$

Combining this with the obvious fact that

$$X^*(t_0) = Z(t_0)$$
$$\psi_0 = \sigma_0^2$$

we see that the coefficients given by (4.1) are just

$$u_i = 1/\psi_i \qquad w_i = 1/\phi_i$$

and (4.2) and (4.3) are equivalent.

This recursive procedure is the idea behind <u>Kalman-filtering</u>, cf. Kalman and Bucy (1961).

<u>Part II</u>. As a result of the first recursion we obtain for each value of i the best predictor of $X(t_i)$ based on $Z(t_0),\ldots,Z(t_i)$. We shall now perform a backwards recursion calculating $\hat{X}(t_i)$ from these values as

$$\left.\begin{aligned} \hat{X}(t_n) &= X^*(t_n) \\ (\omega_{i+1}^{-2} + u_i)\hat{X}(t_i) &= \omega_{i+1}^{-2}\hat{X}(t_{i+1}) + u_i X^*(t_i) \end{aligned}\right\} \tag{4.4}$$

That this is correct is again shown by an induction argument.

Again the form of (4.4) indicates that a heuristic argument of the same kind as in part I can be given although it gets slightly more complicated.

Thiele now proceeds to give recursive procedures for calculating variances of the prediction errors. This is again done elegantly using the following important argument.

It follows from (4.4) that $X^*(t_i)$ are best estimates of the quantities

$$\left.\begin{aligned} A(t_n) &= X(t_n) \\ A(t_i) &= u_i^{-1}\omega_{i+1}^{-2}(X(t_i) - X(t_{i+1})) + X(t_i),\ i < n\ . \end{aligned}\right\} \quad (4.5)$$

The variance of the corresponding errors is thus given as

$$\gamma_n = V(X^*(t_n) - A(t_n)) = u_n^{-1}$$

and for $i < n$

$$\begin{aligned} \gamma_i &= V(X^*(t_i) - A(t_i)) \\ &= u_i^{-2}\ \omega_{i+1}^{-4}\ V(X(t_i) - X(t_{i+1})) + V(X^*(t_i) - X(t_i)) \\ &= u_i^{-2}\ \omega_{i+1}^{-2} + u_i^{-1} = (u_i - w_{i+1})^{-1}\ . \end{aligned}$$

Further, the errors

$$X^*(t_i) - A(t_i),\ i = 0,1,\dots,n$$

are <u>independent</u>. This is typical for the way we have solved the linear equations. To calculate the variance of any prediction error of the type

$$V(\sum_{i=0}^{n} \lambda_i[\hat{X}(t_i) - X(t_i)])$$

we just have to express the linear combination in terms of $X^*(t_i)$

such that

$$\sum_{i=0}^{n} \mu_i X^*(t_i) = \sum_{i=0}^{n} \lambda_i \hat{X}(t_i)$$

and thus also

$$\sum_{i=0}^{n} \mu_i A(t_i) = \sum_{i=0}^{n} \lambda_i X(t_i)$$

whereby

$$V\left(\sum_{i=0}^{n} \lambda_i [\hat{X}(t_i) - X(t_i)]\right) = \sum_{i=0}^{n} \mu_i^2 \gamma_i \ . \tag{4.6}$$

He now attributes to personal communication with Professor Oppermann that it always will be so that as a result of solving the normal equations by the procedure given, one will end up with a system of functions that can be considered as <u>independent</u> and replacing the original observations, thus making the calculation of error variances etc. simple. Some years later Thiele has this idea spelled out systematically by deriving the canonical form of the linear normal model and introducing the notion of a system of "free functions" which is what today is called finding an orthonormal basis of a suitable type, cf. Hald (1981).

Using (4.6) Thiele gives now a recursion for calculating.

$$\left.\begin{aligned} \delta_i^2 &= V(\hat{X}(t_i) - \hat{X}(t_{i-1}) - (X(t_i) - X(t_{i-1}))^2 \\ \tau_i^2 &= V(\hat{X}(t_i) - X(t_i)) \end{aligned}\right\} \tag{4.7}$$

where it for later purpose should be noted that δ_i^2 and τ_i^2 both depend on the entire set of values $\sigma_0^2,\ldots,\sigma_n^2,\omega_1^2,\ldots,\omega_n^2$ but not on α.

Finally it seems worth mentioning that Thiele gives a <u>continued fraction</u> representation of the solutions to the normal equations

(3.2) as well as a description of how to obtain the values $\hat{X}(t_i)$ by <u>geometrical construction</u>.

5. Thiele's estimation of the error variances

The estimation (prediction) described in the previous sections is based on the assumption that the error variances are completely known. Normally, however, the variances are unknown but it will then often be of interest to consider the case

$$\sigma_i^2 = \sigma^2/h_i \qquad \omega_i^2 = \omega^2/k_i \ ,$$

where $h_i > 0$, $i = 0,\dots,n$ and $k_i > 0$, $i = 1,\dots,n$ are known and σ^2, ω^2 unknown. A typical situation could be

$$h_i = 1 \ , \quad k_i = (t_i - t_{i-1})^{-1} \ .$$

We are then faced with the problem of estimating σ^2 and ω^2. Thiele discusses this problem and gives a heuristic argument for his solution. We shall describe his method and argument and investigate it from a more exact point of view.

First Thiele claims that it seems appropriate to base an estimate of σ^2 and ω^2 on the quadratic forms

$$Q_1 = \sum_{i=0}^{n} h_i (Z(t_i) - \hat{X}(t_i))^2$$

and

$$Q_2 = \sum_{i=1}^{n} k_i (\hat{X}(t_i) - \hat{X}(t_{i-1}))^2$$

where $\hat{X}(t_0),\dots,\hat{X}(t_n)$ are calculated as described in the previous section from certain <u>initial values</u> σ_0^2, ω_0^2 of σ^2 and ω^2.

The first problem is to decide how the degrees of freedom should be allocated to the two quadratic forms. The total number of degrees of freedom must be n since we have used the method of least squares on 2n+1 equations with n+1 unknowns. It seems thus appropriate to let

$$\tilde{\sigma}^2 = Q_1/f_1 \ ; \quad \tilde{\omega}^2 = Q_2/f_2 \tag{5.1}$$

where $f_1 + f_2 = n$ and (f_1, f_2) are chosen in a reasonable way.

In analogy with usual least squares where only one variance has to be estimated it seems plausible that one should subtract from the number of terms, the relative amount of variation due to the estimation of $X(t_i)$. More precisely let

$$f_1 = n + 1 - \sigma_0^{-2} \sum_{i=0}^{n} h_i \tau_i^2(\sigma_0^2, \omega_0^2)$$
$$f_2 = n \qquad - \omega_0^{-2} \sum_{i=1}^{n} k_i \delta_i^2(\sigma_0^2, \omega_0^2)$$

where $\tau_i^2(\sigma_0^2, \omega_0^2)$ and $\delta_i^2(\sigma_0^2, \omega_0^2)$ are the prediction errors in (4.7) based on the assumption that $\sigma_i^2 = \sigma_0^2/k_i$ and $\omega_i^2 = \omega_0^2/k_i$. Note that in fact

$$f_1 + f_2 = n$$

since

$$\sigma_0^{-2} \sum_{i=0}^{n} h_i \tau_i^2(\sigma_0^2, \omega_0^2) + \omega_0^{-2} \sum_{i=1}^{n} k_i \delta_i^2(\sigma_0^2, \omega_0^2)$$

is the trace of the matrix of a projection onto an n+1 dimensional subspace of $\mathbb{R}^{2n+1}$ and thus equal to n+1.

Using (5.1) we obtain new values $\tilde{\sigma}^2$ and $\tilde{\omega}^2$ of σ^2 and ω^2 and we then repeat the procedure in the sense that new estimates $\hat{X}(t_i)$ are calculated, new values for τ_i^2, δ_i^2, Q_1, Q_2, f_1 and f_2 etc. The procedure is to be repeated until stable values of σ^2, ω^2 are reached. According to Thiele one has to do that three or four times. The final stable values of $\tilde{\sigma}^2$, $\tilde{\omega}^2$ are then used as estimates of σ^2 and ω^2. Thiele writes that "it seems at least plausible" that the procedure is correct.

Formulating Thiele's estimation method more precisely we see that his estimates $\hat{X}(t_0),\dots,\hat{X}(t_n)$, $\tilde{\sigma}^2$, $\tilde{\omega}^2$ satisfy the system of $n+3$ equations obtained by taking (3.2), inserting into these the values

$$\sigma_i^2 = \tilde{\sigma}^2/h_i \qquad \omega_i^2 = \tilde{\omega}^2/k_i$$

and supplementing with the equations

$$\left.\begin{aligned} \tilde{\sigma}^2 &= \frac{\sum_{i=0}^{n} h_i (Z(t_i) - \hat{X}(t_i))^2}{n+1-\tilde{\sigma}^{-2}\sum_{i=0}^{n} h_i \tau_i^2(\tilde{\sigma}^2,\tilde{\omega}^2)} \\ \tilde{\omega}^2 &= \frac{\sum_{i=1}^{n} k_i (\hat{X}(t_i) - \hat{X}(t_{i-1}))^2}{n-\tilde{\omega}^{-2}\sum_{i=1}^{n} k_i \delta_i^2(\sigma^2,\omega^2)} \quad . \end{aligned}\right\} \tag{5.3}$$

Rearranging (5.3) we get

$$\left.\begin{aligned} (n+1)\tilde{\sigma}^2 &= \sum_{i=1}^{n} h_i (Z(t_i) - \hat{X}(t_i))^2 + \sum_{i=0}^{n} h_i E_{\tilde{\sigma}^2,\tilde{\omega}^2}(X(t_i) - \hat{X}(t_i))^2 \\ n\tilde{\omega}^2 &= \sum_{i=1}^{n} k_i (\hat{X}(t_i) - \hat{X}(t_{i-1}))^2 + \sum_{i=0}^{n} k_i E_{\sigma^2,\omega^2}(\Delta_i)^2 \end{aligned}\right\} \tag{5.4}$$

where

$$\Delta_i = \hat{X}(t_i) - \hat{X}(t_{i-1}) - (X(t_i) - X(t_{i-1})) \quad .$$

The form (5.4) will be convenient in the next section.

Thiele does not discuss the problems of existence and uniqueness of solutions to equations (5.4) and (3.2) combined, nor does he discuss convergence properties of the iterative procedure to solve these equations beyond the remarks mentioned earlier that it has to be performed "three or four times". We shall return to this problem later.

6. Discussion of procedures for estimation of the error variances

We shall first reformulate the problem slightly by forgetting the prediction problem and consider the problem of estimating the unknown values of $\alpha, \sigma^2, \omega^2$ based on the observations

$$Z(t_0) = z_0, \ldots, Z(t_n) = z_n$$

where $(Z(t_0), \ldots, Z(t_n))$ has the joint normal distribution specified by Thiele's model.

We shall treat the problem as a problem of estimation in an exponential family with <u>incomplete observation</u>, cf. Sundberg (1974).

That is, we shall use that we can think of $Z(t_i)$ as

$$Z(t_i) = \alpha + Y(t_i) + \varepsilon(t_i)$$

where $Y(t) = X(t) - \alpha$ is a Gaussian process with expectation equal to zero, independent increments with

$$Y(t_i) - Y(t_{i-1}) = X(t_i) - X(t_{i-1}) \sim N(0, \omega^2/k_i) \quad ,$$

$Y(t_0) = 0$ and $\varepsilon(t_i)$ independent and independent of the Y's with expectation equal to zero and variance $\sigma_i^2 = \sigma^2/h_i$.

Suppose that we had observed not only $Z(t_i) = z_i$, $i = 0, \ldots, n$ <u>but also</u> $Y(t_i) = y_i$ for $i = 0, \ldots, n$, where $y_0 = 0$. The likelihood of α, β, σ^2 would then be given as

$$
\begin{aligned}
&- 2\log L(\alpha,\sigma^2,\omega^2) \\
&= \psi(\alpha,\sigma^2,\omega^2) + \sigma^{-2}\sum_{i=0}^{n} h_i(z_i - y_i - \alpha)^2 + \omega^{-2}\sum_{i=1}^{n} k_i(y_i - y_{i-1})^2 \\
&= \psi^*(\alpha,\sigma^2,\omega^2) + \sigma^{-2}\sum_{i=0}^{n} h_i(z_i - y_i)^2 + \omega^{-2}\sum_{i=1}^{n} k_i(y_i - y_{i-1})^2 \\
&\qquad - 2\alpha\sigma^{-2}\sum_{i=0}^{n} h_i(z_i - y_i) \ .
\end{aligned}
$$

Thus we see that we deal with an exponential family with canonical statistics

$$
\begin{aligned}
s_1 &= \sum_{i=0}^{n} h_i(z_i - y_i)^2 \\
s_2 &= \sum_{i=1}^{n} k_i(y_i - y_{i-1})^2 \\
s_3 &= \sum_{i=0}^{n} h_i(z_i - y_i) \ .
\end{aligned}
$$

From general theory, Sundberg (1974), it now follows that we get the likelihood equations <u>in the case where only $\underset{\sim}{z}$ has been observed</u> by equating the expectation of these statistics to their conditional expectations given the observed value of $\underset{\sim}{z}$.

We get for the expectations

$$
\left.
\begin{aligned}
ES_1 &= (n+1)\sigma^2 + \alpha^2\sum_{i=0}^{n} h_i \\
ES_2 &= n\omega^2 \\
ES_3 &= \alpha\sum_{i=0}^{n} h_i
\end{aligned}
\right\} \qquad (6.1)
$$

and the conditional expectations need to be worked upon a bit:

$$
\begin{aligned}
&E(S_1 \mid Z(t_0) = z_0,\dots,Z(t_n) = z_n) \\
&\quad = E(S_1 \mid \underset{\sim}{z}) = \sum_{i=0}^{n} h_i\, E((Z(t_i) - Y(t_i))^2 \mid \underset{\sim}{z}) \\
&\quad = \sum_{i=0}^{n} h_i(z_i - \hat{Y}(t_i))^2 + \sum_{i=0}^{n} h_i\, E((\hat{Y}(t_i) - Y(t_i))^2 \mid \underset{\sim}{z}) \qquad (6.2)
\end{aligned}
$$

where we have let

$$\hat{Y}(t_i) = E_\alpha(Y(t_i) \mid z) = X^\alpha(t_i) - \alpha$$

where

$$X^\alpha(t_i) = E_\alpha(X(t_i) \mid \underset{\sim}{z}) ,$$

such that Thiele's $\hat{X}(t_i)$-values are given as

$$\hat{X}(t_i) = \hat{X^\alpha}(t_i) \tag{6.3}$$

The term $E((\hat{Y}(t_i) - Y(t_i))^2 \mid \underset{\sim}{z})$ is just a conditional variance and does therefore not depend on $\underset{\sim}{z}$ so that we can proceed as

$$\begin{aligned} E(S_1 \mid \underset{\sim}{z}) = \sum_{i=0}^{n} h_i (z_i - X^\alpha(t_i))^2 + \alpha^2 \sum_{i=0}^{n} h_i \\ + 2\alpha \sum_{i=0}^{n} h_i (z_i - X^\alpha(t_i)) \\ + \sum_{i=0}^{n} h_i E(X^\alpha(t_i) - X(t_i))^2 \quad . \end{aligned} \tag{6.4}$$

Similarly we get for S_2 :

$$\begin{aligned} E(S_2 \mid \underset{\sim}{z}) &= \sum_{i=1}^{n} k_i \; E((Y(t_i) - Y(t_{i-1}))^2 \mid \underset{\sim}{z}) \\ &= \sum_{i=1}^{n} k_i \; E((\hat{Y}(t_i) - \hat{Y}(t_{i-1}))^2 \mid \underset{\sim}{z}) \\ &\quad + \sum_{i=1}^{n} k_i \; E((\hat{Y}(t_i) - \hat{Y}(t_{i-1}) - (Y(t_i) - Y(t_{i-1})))^2 \mid \underset{\sim}{z}) \\ &= \sum_{i=1}^{n} k_i (\hat{Y}(t_i) - \hat{Y}(t_{i-1}))^2 \\ &\quad + \sum_{i=1}^{n} k_i \; E(\hat{Y}(t_i) - \hat{Y}(t_{i-1}) - (Y(t_i) - Y(t_{i-1})))^2 \\ &= \sum_{i=1}^{n} k_i (X^\alpha(t_i) - X^\alpha(t_{i-1}))^2 \\ &\quad + \sum_{i=1}^{n} k_i \; E(X^\alpha(t_i) - X^\alpha(t_{i-1}) - (X(t_i) - X(t_{i-1})))^2 , \end{aligned} \tag{6.5}$$

and finally for S_3:

$$E(S_3|\underset{\sim}{z}) = \sum_{i=0}^{n} h_i(z_i - \hat{Y}(t_i))^2$$

$$= \sum_{i=0}^{n} h_i(z_i - X^{\alpha}(t_i)) + \alpha \sum_{i=0}^{n} h_i \quad . \tag{6.6}$$

We now form the equations obtained by equating (6.1) to (6.4), (6.5) and (6.6) and get from the last of these that the solution $\hat{\alpha}$ satisfies

$$\sum_{i=0}^{n} h_i(z_i - X^{\hat{\alpha}}(t_i)) = 0 \quad .$$

Inserting this into the two first equations gives us the equations

$$(n+1)\hat{\sigma}^2 = \sum_{i=0}^{n} h_i(z_i - X^{\hat{\alpha}}(t_i)) + \sum_{i=0}^{n} h_i E_{\hat{\sigma}^2,\hat{\omega}^2}(X^{\alpha}(t_i) - X(t_i))^2$$

$$n\hat{\omega}^2 = \sum_{i=1}^{n} k_i(X^{\hat{\alpha}}(t_i) - X^{\hat{\alpha}}(t_i))^2$$

$$+ \sum_{i=1}^{n} k_i E_{\hat{\sigma}^2,\hat{\omega}^2}(X^{\alpha}(t_i) - X^{\alpha}(t_{i-1}) - (X(t_i) - X(t_{i-1})))^2 ,$$

(6.7)

where we have made clear how the expectations depend on the parameters. To see that these equations are very similar to Thiele's (5.4) we use (6.3) and the fact that

$$E_{\sigma^2,\omega^2}(\hat{X}(t_i) - X(t_i))^2 = E_{\sigma^2,\omega^2}(X^{\alpha}(t_i) - X(t_i) + \hat{\alpha} - \alpha)^2$$

$$= E_{\sigma^2,\omega^2}(X^{\alpha}(t_i) - X(t_i))^2 + E_{\alpha^2,\omega^2}(\hat{\alpha} - \alpha)^2$$

and

$$E_{\sigma^2,\omega^2}(X^{\alpha}(t_i) - X^{\alpha}(t_{i-1}) - (X(t_i) - X(t_{i-1})))^2 = E_{\sigma^2,\omega^2}(\Delta_i)^2$$

whereby we see that the only difference between Thiele's equations (5.4) and the maximum likelihood equations (6.7) is the term

$$\sum_{i=1}^{n} h_i \, E(\hat{\alpha} - \alpha)^2$$

on the right hand side of the first equation.

In the limiting case with $k_i \equiv \infty$, i.e. where the Brownian motion is vanishing and $Z(t_0),\ldots,Z(t_n)$ become independent we have

$$\hat{X}^{\alpha}(t_i) = \hat{X}(t_i) = \hat{\alpha} = \left(\sum_{i=0}^{n} h_i\right)^{-1} \sum_{i=0}^{n} h_i z_i$$

$$E_{\sigma^2,\omega^2}(\hat{\alpha} - \alpha)^2 = \left(\sum_{i=0}^{n} h_i\right)^{-1} \sigma^2$$

whereby (6.7) reduces to the equation

$$(n+1)\sigma^2 = \sum_{i=0}^{n} h_i (z_i - \hat{\alpha})^2 \qquad (6.8)$$

whereas (5.4) becomes

$$(n+1)\tilde{\sigma}^2 = \sum_{i=0}^{n} h_i (z_i - \hat{\tilde{\alpha}})^2 + \tilde{\sigma}^2 \qquad (6.9)$$

so that Thiele's equations correspond to the usual way of taking into account that a linear parameter has been estimated.

Thiele's equations (5.4) can be seen to be equivalent to the equations defining the so-called <u>restricted maximum likelihood estimates</u>, cf. Patterson and Thompson (1971). The equations can e.g. be obtained as follows. Write the joint density of $Z(t_0),\ldots,Z(t_n)$ as

$$f(z_0,\ldots,z_n) = h(z_0 | z_1 - z_0,\ldots,z_n - z_0)\, g(z_1 - z_0,\ldots,z_n - z_0)$$

where g is the marginal density of the n linear "contrasts" and h is the conditional density of z_0 given these contrasts(any set of $n+1$ linearly independent linear combinations of which the n are contrasts will do).

The likelihood equations for σ^2, ω^2 based entirely on the marginal distribution of the contrasts is exactly Thiele's equations (5.4), see the cited paper by Patterson and Thompson (1971) and also Harville (1977).

Also the algorithm used by Thiele is identical to the algorithm suggested by Patterson and Thompson (1971).

Harville (1977) has discussed the properties of the algorithm although on a heuristic and empirical basis. Since the algorithm is a special case of the more general EM-algorithm discussed by Dempster, Laird and Rubin (1977) one can say a bit more about its properties.

If the starting values do not happen to be a saddle point of the likelihood function, it either converges to a local maximum of the likelihood function or diverges in the sense that either $\sigma^2 \to 0$ or $\omega^2 \to 0$. Each step of the algorithm increases the likelihood. There seem to be no simple conditions for the equations to have a unique solution.

Thiele seemed unaware of these problems not even mentioning them. A guess would be that his long series of observations ensured that he never encountered the problem in practice. Also it seems as if he did not pay much attention at all to the variance estimation and only wanted to correct the initial values of (σ^2, ω^2) as far as this gave a significant change in the values of $\hat{X}(t_i)$

that were of primary interest to him. It is anyway quite typical for the time period that variances only have secondary importance.

The last section of Thiele's paper contains a worked out example with a series of 74 astronomical measurements and here he has no problems.

7. An application

In the final part of Thiele's paper he extends the model by the inclusion of a linear regression term in the sense that he considers the problem

$$Z(t_i) = \sum_{\nu=1}^{k} \alpha_\nu f_\nu(t_i) + X(t_i) + \varepsilon(t_i)$$

where $f_1, \ldots, f_k$ are known functions and $\alpha_1, \ldots, \alpha_k$ are unknown constants. He gives a similar recursive procedure for this problem although he leaves the proofs to the reader.

Working with hormone concentrations in pregnancy some time ago (Lauritzen (1976)) I used the model of the form

$$Z(t_i) = \alpha + \beta t_i + X(t_i)$$

for the logarithm of concentrations of progesterone in plasma. The data showed good fit but indicated that a model of the type

$$Z(t_i) = \alpha + \beta t_i + X(t_i) + \varepsilon(t_i)$$

would be more realistic and give a better fit, although the quality and amount of data prevented to pursue the issue further.

Recently Mogens Christensen, Aalborg Hospital, has provided me with measurements of concentrations of the hormone HPL on 69 pregnant women, taken at various time points during pregnancy and with 4-11 observations for each women. It seemed natural to try out Thiele's model and algorithm on this data set. Of course one could expect difficulties because of the very short observational series and this was indeed the case. The algorithm did <u>not</u> converge for 54 of the 69 series. Thus it looks as if the distribution of the solutions to the estimating equations have a lot of mass on the boundary and the only way to get around this was to assume the variances of the Brownian motion and the white noise to be identical for all women. Under this assumption the algorithm converged in the sense that after 24 iterations the estimates for the variances did not change with more than 1% and after 42 iterations not more than 1 o/oo. The stable values of σ^2, ω^2 reached were quite sensible compared to other knowledge and examination of residuals showed a good fit. On the other hand 42 iterations is a lot more than "three or four times" as Thiele writes. Nowadays these iterations can be performed quickly and cheaply on a high speed computer, but in 1880... ?

Note that it in this particular application is quite important that one uses Thiele's 'restricted maximum likelihood estimates' rather than the maximum likelihood estimates themselves since otherwise the different values of α and β for different women would create a nuisance parameter effect and give rise to useless estimates of σ^2 and ω^2.

8. Final comments

Even though Thiele did not fully discuss the difficulties concerning the estimation of the error variances, it is amazing that he got so far in the understanding of the time series model that he discussed.

The most striking contrast between Thiele's "approach" to time series analysis and time series analysis as of today seems to be the very detailed analysis of one particular model, as opposed to the more modern tendency of investigating large classes of models without really trying to understand or utilise the particular structure of each of them. Also the way Thiele establishes his model is closely related to a particular practical problem where he takes very much into account, how the observations in fact have been produced. Sometimes one can get the impression that model building in time series analysis today is made completely independent of how the data have been generated.

Could Thiele's success in 1880 encourage modern statisticians to make detailed investigations of particular models ?

Acknowledgements I wish to thank A. Hald for several discussions on the topics in the present paper and for reading the manuscript and suggesting several important improvements.

Literature

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J.Roy.Statist.Soc. B, **39**, 1-22.

Hald, A. (1981). T.N. Thiele's contributions to statistics. Int.Statist.Review, **49** (to appear).

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J.Amer.Statist.Ass., **72**, 320-338.

Helmert, F.R. (1907). Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. 2nd.ed., Teubner, Leipzig.

Kalman, R.E. and Bucy, R. (1961). New results in linear filtering and prediction. J.Basic Engineering, **83**D, 95-108.

Lauritzen, S.L. (1976). Appendix to Winkel et.al: Method for monitoring plasma progesterone concentrations in pregnancy. Clinical Chemistry, **22**, 427-428.

Patterson, H.D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. Biometrika, **58**, 545-554.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications. 2nd.ed., Wiley, New York.

Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. Scand.J.Statist., **1**, 49-58.

Thiele, T.N. (1880). Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en "systematisk" Karakter. Vidensk.Selsk.Skr. 5.rk., naturvid. og mat. Afd., **12**, 5, 381-408.

Thiele, T.N. (1880a). Sur la compensation de quelques erreurs quasi-systematiques par la méthode des moindres carreés. Reitzel, København. (French translation of Thiele (1880)).

PREPRINTS 1980

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Olkin, Ingram and Væth, Michael: Maximum Likelihood Estimation in a Two-way Analysis of Variance with Correlated Errors in One Classification.

No. 2 Skovgaard, Ib M.: Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions.

No. 3 Asmussen, Søren: Equilibrium Properties of the M/G/l Queue.

No. 4 Johansen, Søren and Keiding, Susanne: A Family of Models for the Elimination of Substrate in the Liver.

No. 5 Skovgaard, Ib M.: Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators.

No. 6 Tjur, Tue: A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model.

No. 7 Lauritzen, Steffen L. : Examples of Extreme Point Models in Statistice.

PREPRINTS 1981

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen L.: Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.