Maximum Likelihood Estimation in Forest Growth Models with Measurement Errors

Alex Teterukovskiy and Bo Ranneby

Research Report
Centre of Biostochastics

Swedish University of
Agricultural Sciences

Report 2003:3
ISSN 1651-8543

Maximum Likelihood Estimation in Forest Growth Models with Measurement Errors

Alex Teterukovskiy and Bo Ranneby

Centre of Biostochastics
Swedish University of Agricultural Sciences
SE-90183, Umeå, Sweden

Abstract

The forest growth model with measurement errors is introduced. The maximum likelihood estimates (MLE) of the parameters of this model are proven to be consistent and asymptotically normally distributed. The model is applied to the real data from Swedish National Forest Inventory and the MLE of the parameters are obtained.

Keywords: Forest growth model; Maximum likelihood estimation; Random optimization

1 Introduction

This study was motivated by the need for accurate and applicable single-tree growth functions, that exists in modern forestry. The growth of the tree apparently depends on the condition of the tree itself, but also on the tree's competitive situation and on the conditions of the site. Such functions have been developed in many countries at least during the past thirty years (Stage (1974), Ek and Monserud (1974), Siitoen (1995)). In practice, methods of regression analysis are often used to predict the forest growth (see, for example, Candy (1997) and Elfving (2000)). The dependent variable in the regression is usually the increase in volume or basal area, which are calculated based on measurements on two occasions. It is clear that this value is highly influenced by measurement errors. Therefore we have introduced a model that takes into account the measurement errors, producing more accurate estimates of the parameters. The model is described in Section 2. The properties of the maximum likelihood estimates of the parameters can be established using the results by Hoadley (1971). We derive the asymptotic properties of the maximum likelihood estimates (MLE) of the parameters of the model in Section 3, and, finally, in Section 4 we apply our model to the real data taken from the Swedish National Forest Inventory. We compare the results to those obtained by means of a standard model, and draw final conclusions in favor of the model proposed in the current paper.

2 Model

The well-known forest growth linear regression model is

$$\ln Y_k = \boldsymbol{\beta}^T \mathbf{x}_k + U_k, \tag{1}$$

where Y_k is the increase in squared diameter (or basal area) and $\mathbf{x}_k$ is a vector of size p of explanatory variables associated with a tree at location k, $\boldsymbol{\beta}$ is a vector of regression coefficients (including the intercept) and U_k is the deviation from the regression function. The Y_k are assumed to be independent and possessing the same variance.

Unfortunately, it often happens that because of the measurement errors, the measured increase in basal area becomes negative. It is also true that the measurement errors differ from tree to tree, increasing with the diameter. In order to cope with the mentioned complications, in Elfving (2000) the model (1) was substituted by the non-linear weighted variant:

$$Y_k = e^{\boldsymbol{\beta}^T \mathbf{x}_k} + V_k, \tag{2}$$

where V_k is the normal deviation from the regression function. For details of the weighting procedure, we refer to the original paper. Comparing models (1) and (2), the following discrepancy becomes evident - the variation of growth is expected to have a log-normal distribution, whereas the distribution of the

measurement error is expected to be normal. Model (2) oversimplifies this issue by assuming the normal residual distribution, and therefore, should be considered inappropriate.

As an alternative to (2) and a natural extension of (1), we suggest the following model:

$$Y_k = e^{\boldsymbol{\beta}^T \mathbf{x}_k + U_k} + W_k, \tag{3}$$

where W_k is a normally distributed measurement error, whose variance depends through the known positive function g upon vector $\mathbf{x}_k$:

$$W_k \sim N(0, \tau^2 g(\mathbf{x}_k)). \tag{4}$$

The proportion coefficient τ^2 in (4) is to be estimated.

Suppose that the distribution of the deviation U_k does not depend on k and is Gaussian $N(\mu, \sigma^2)$. Without loss of generality we may adjust the intercept in the model and demand that the parameters obey the following condition of unbiasedness:

$$E e^{\boldsymbol{\beta}^{\mathbf{T}} \mathbf{x}_k + U_k} = e^{\boldsymbol{\beta}^T \mathbf{x}_k}.$$

From this relation we obtain $\mu = -\sigma^2/2$, which reduces by one the number of unknown parameters. Therefore, the total number of parameters to be estimated is $m = p + 2$.

It is clear that (3) is a non-linear regression model with an error term distributed according to the convolution of the lognormal and normal distributions (LNN).

The density function of the LNN can be seen to be

$$f_k(y_k) = \int_{-\infty}^{\infty} A_k(t) dt, \tag{5}$$

where

$$A_k(t) = \frac{1}{2\pi\sigma\sqrt{g(\mathbf{x}_k)}\tau} e^{-\frac{1}{2}\left(\frac{(t - \sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{\sigma^2} + \frac{(y_k - e^t)^2}{g(\mathbf{x}_k)\tau^2} \right)}.$$

Although the density function of LNN takes form of a non-standard integral, some attractive properties of this distribution can be easily established. The density function of LNN has no singularities, is infinitely differentiable with respect to all its parameters, and, finally, the LNN distribution possesses all moments (see, for example, Hawkins (1991)).

We suppose that the parameter space is compact, and σ and τ are bounded away from zero, e.g. $\sigma \geq \delta > 0$, $\tau \geq \delta > 0$. We also assume that the values of $\mathbf{x}_k$ are bounded (this is always true in applications).

Let us denote the vector of the parameters by $\boldsymbol{\theta}$=$\{\theta_1, \theta_2, \cdots, \theta_m\}$. For a sequence of independent random variables $Y_k, k = 1, 2, \ldots$ which take values in

some measurable space and possess densities $f_k(y_k, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$, the likelihood function takes form

$$L_n(\boldsymbol{\theta}) = \prod_{k=1}^{n} f_k(y_k, \boldsymbol{\theta}).$$

The MLE of the true parameter value $\boldsymbol{\theta}^0$ is denoted by $\hat{\boldsymbol{\theta}}_n$ and is defined to be any point in Θ such that

$$L_n(\hat{\boldsymbol{\theta}}_n) \geq L_n(\boldsymbol{\theta}) \text{ for all } \boldsymbol{\theta} \in \Theta.$$

3 Properties of MLE

In order to establish the consistency and the asymptotic normality of the MLE for model (3), we will use two theorems due to Hoadley (1971).

For convenience let us define the following variables:

$$R_k(\boldsymbol{\theta}) = \log \frac{f_k(y_k, \boldsymbol{\theta})}{f_k(y_k, \boldsymbol{\theta}^0)},$$

$$R_k(\boldsymbol{\theta}, \rho) = \sup_{\|t - \boldsymbol{\theta}\| \leq \rho} R_k(t),$$

$$V_k(r) = \sup_{\|\boldsymbol{\theta}\| \geq r} R_k(\boldsymbol{\theta}).$$

For any random variable X, let

$$\begin{aligned} X^{(B)} &= X \quad \text{if } X \geq -B \\ &= -B \quad \text{otherwise,} \end{aligned}$$

where $B \geq 0$. Let the positive constants K and δ be generic, e.g. a bound involving them must hold for $k = 1, 2, \ldots$. Finally, define $\Phi_k(Y_k, \boldsymbol{\theta}) = \log f_k(Y_k, \boldsymbol{\theta})$.

Theorem 1 below is from Hoadley (1971) and is essentially an extension to the independent non-identically distributed case of a result from Wald (1949).

3.1 Consistency

THEOREM 1. *Consider the following assumptions:*

I Θ *is a closed subset of* $\mathbb{R}^m$.

II $f_k(y_k, \boldsymbol{\theta})$ *is an upper semi-continuous function of* $\boldsymbol{\theta}$, *uniformly in* k, *a.s.*

III *There exist* $\rho^* = \rho^*(\boldsymbol{\theta}) > 0$ *and* $r > 0$ *for which*

i $E[R_k(\boldsymbol{\theta}, \rho)]^{1+\delta} \leq K, 0 \leq \rho \leq \rho^*$;

ii $E[V_k(r)]^{1+\delta} \leq K$.

IV *There exists* $B > 0$ *for which*

i $\limsup_{n \longrightarrow \infty} \frac{1}{n} \sum E_{\boldsymbol{\theta}^0} R_k^{(B)}(\boldsymbol{\theta}) < 0, \boldsymbol{\theta} \neq \boldsymbol{\theta}^0$;

ii $\limsup_{n \longrightarrow \infty} \frac{1}{n} \sum E_{\boldsymbol{\theta}^0} V_k^{(B)}(r) < 0$.

V $R_k(\boldsymbol{\theta}, \rho)$ *and* $V_k(r)$ *are measurable functions of* y_k.

If Assumptions ***I-V*** *are satisfied, then, if* $\hat{\boldsymbol{\theta}}_n$ *represents an MLE for* $\boldsymbol{\theta}$ *and* $\boldsymbol{\theta}^0$ *is the true value of* $\boldsymbol{\theta}$,

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^0.$$

Apparently, Assumption **I** holds. Assumptions **II** and **V** are then also obviously satisfied.

To verify Assumption **III** we first obtain the following bound for $f_k(y_k)$:

$$\begin{aligned}
f_k(y_k) &= C_1(\boldsymbol{\theta}, \mathbf{x}_k) e^{-\frac{y_k^2}{2g(\mathbf{x}_k)\tau^2}} \int_{-\infty}^{\infty} e^{-\frac{(t-\sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{2\sigma^2}} e^{\frac{y_k e^t}{g(\mathbf{x}_k)\tau^2}} e^{-\frac{e^{2t}}{2g(\mathbf{x}_k)\tau^2}} dt \\
&= C_1(\boldsymbol{\theta}, \mathbf{x}_k) e^{-\frac{y_k^2}{2g(\mathbf{x}_k)\tau^2}} \int_{-\infty}^{\infty} e^{-\frac{(-t-\sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{2\sigma^2}} e^{\frac{y_k e^{-t}}{g(\mathbf{x}_k)\tau^2}} e^{-\frac{e^{-2t}}{2g(\mathbf{x}_k)\tau^2}} dt \\
&\geq C_1(\boldsymbol{\theta}, \mathbf{x}_k) e^{-\frac{y_k^2}{2g(\mathbf{x}_k)\tau^2}} \int_{t}^{\infty} e^{-\frac{(-t-\sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{2\sigma^2}} \min(1, e^{\frac{y_k}{g(\mathbf{x}_k)\tau^2}}) e^{-\frac{1}{2g(\mathbf{x}_k)\tau^2}} dt \\
&= C_2(\boldsymbol{\theta}, \mathbf{x}_k) e^{-\frac{y_k^2}{2g(\mathbf{x}_k)\tau^2}} \min(1, e^{\frac{y_k}{g(\mathbf{x}_k)\tau^2}}),
\end{aligned}$$

where $C_i(\boldsymbol{\theta}, \mathbf{x}_k), i = 1, 2$, are positive continuous functions.

On the other hand, f_k is obviously bounded from above. Therefore, taking into account that the LNN distribution possesses all moments, it can be inferred that Assumption **III** holds for some K (if we choose $\delta = 1$, existence of moments of order 4 is sufficient).

From Jensen's inequality it follows that if $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$, then $E_{\boldsymbol{\theta}^0} R_k(\boldsymbol{\theta}) < 0$. In our case, $E_{\boldsymbol{\theta}^0} R_k(\boldsymbol{\theta})$ is a negative continuous bounded function of $\mathbf{x}_k$. Taking into account that the values of $\mathbf{x}_k$ are bounded, it follows that $E_{\boldsymbol{\theta}^0} R_k^{(B)}(\boldsymbol{\theta}) < -r < 0$ for some $r, B > 0$ and for all k. The same considerations are valid for $E_{\boldsymbol{\theta}^0} V_k^{(B)}(r)$. We conclude that Assumption **IV** holds.

3.2 Asymptotic normality

Now, when we have established the consistency of the MLE, we will verify the assumptions sufficient to establish the asymptotic normality of the MLE. We use Theorem 2 from Hoadley (1971) which is an extension of the result obtained in Roussas (1968).

THEOREM 2. *Consider the following assumptions:*

I Θ *is an open subset of* $\mathbb{R}^m$.

II $\hat{\boldsymbol{\theta}}_n$ *converges in probability to* $\boldsymbol{\theta}^0$.

III $\dot{\Phi}_k(Y_k, \boldsymbol{\theta})$ *and* $\ddot{\Phi}_k(Y_k, \boldsymbol{\theta})$ *exist, a.s.*

IV $\ddot{\Phi}_k(Y_k, \boldsymbol{\theta})$ *is a continuous function of* $\boldsymbol{\theta}$*, uniformly in* k*, a.s., and is a measurable function of* Y_k.

V $E[\dot{\Phi}_k(Y_k, \boldsymbol{\theta})|\boldsymbol{\theta}] = 0, k = 1, 2, \cdots$.

VI $\Gamma_k(\boldsymbol{\theta}) = E[\dot{\Phi}_k(Y_k, \boldsymbol{\theta})\dot{\Phi}_k(Y_k, \boldsymbol{\theta})'|\boldsymbol{\theta}] = -E[\ddot{\Phi}_k(Y_k, \boldsymbol{\theta})|\boldsymbol{\theta}]$.

VII $\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{k=1}^{n} \Gamma_k(\boldsymbol{\theta}) = \bar{\Gamma}(\boldsymbol{\theta})$.

VIII *For some* $\delta > 0$, $\sum_k E|\lambda'\dot{\Phi}_k(Y_k, \boldsymbol{\theta}^0)|^{2+\delta}/n^{(2+\delta)/2} \to 0$ *for all* $\lambda \in \mathbb{R}^m$.

IX *There exist* $\epsilon > 0$ *and random variables* $B_{k,ij}(Y_k)$ *such that*

i $\sup\{|\ddot{\Phi}_{k,ij}(Y_k, t)| : ||t - \boldsymbol{\theta}^0|| \leq \epsilon\} \leq B_{k,ij}(Y_k)$;

ii $E|B_{k,ij}(Y_k)|^{1+\delta} \leq K$.

Given Assumptions ***I-IX****,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{L} N(\mathbf{0}, \bar{\Gamma}^{-1}(\boldsymbol{\theta}^0)).$$

For Assumption **I** we note that since the consistency is established, now it suffices to consider an open neighborhood of the true parameter value. Assumptions **II-VI** are standard, and are obviously satisfied. We will first verify Assumptions **VIII** and **IX**, and then comment on Assumption **VII**.

Using the generalized first mean value theorem, we note that for every i

$$\dot{\Phi}_{k,i}(Y_k, \boldsymbol{\theta}) = \frac{\int_{-\infty}^{\infty} P_2(Y_k, t, e^t)A_k(t)dt}{\int_{-\infty}^{\infty} A_k(t)dt} =$$

$$= \frac{\int_{-\infty}^{\infty} P_2(Y_k, t, e^t)e^{-t}e^{-\frac{(t-\sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{2\sigma^2}} e^t e^{-\frac{(Y_k - e^t)^2}{2g(\mathbf{x}_k)\tau^2}} dt}{\int_{-\infty}^{\infty} e^{-t}e^{-\frac{(t-\sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2}{2\sigma^2}} e^t e^{-\frac{(y_k - e^t)^2}{2g(\mathbf{x}_k)\tau^2}} dt} = P_2^*(Y_k, \boldsymbol{\theta}),$$

Table 1
Initial values and MLE of the parameters

Parameter	Explanatory variable description	Estimates	
		Initial	MLE
β_0	Constant	-0.501	-0.643
β_1	Log(Diameter)	1.091	0.981
β_2	Inverse basal area	10.894	10.437
β_3	Basal area of larger trees	-0.028	-0.027
β_4	Vegetation type (indicator)	0.552	0.304
β_5	Recently thinned (indicator)	0.167	0.352
β_6	Ditch nearby (indicator)	0.132	0.207
β_7	Soil type (indicator)	-0.524	-0.106
β_8	Shallow soil layer (indicator)	-0.215	-0.403
β_9	Divided plot (indicator)	0.103	0.326
β_{10}	Temperature sum	0.620	1.009
β_{11}	Proportion of other species	0.302	0.382
β_{12}	$\text{Log}\frac{\text{basal area on the plot}}{\text{basal area of the surrounding stand}}$	0.428	0.543
σ^2	Variance of the growth model	1	0.628
τ^2	Proportion coefficient of variance of measurement error	0	5.193

where P_2 and P_2^* are some polynomials of order two, both depending on $\boldsymbol{\theta}$. Taking into the account that LNN distribution possesses all moments, it immediately follows that Assumption **VIII** holds. Similar transformations applied to $\ddot{\Phi}_k(Y_k, \boldsymbol{\theta})$ yield

$$\ddot{\Phi}_k(Y_k, \boldsymbol{\theta}) = P_4^*(Y_k, \boldsymbol{\theta}),$$

which implies Assumption **IX**. We have listed the necessary derivatives of the LNN likelihood function in the Appendix.

Assumption **VII** is not so straightforward to verify. Indeed, it is possible to show that the positive definite limit does not have to exist for arbitrary set of values of explanatory variables. However, such examples will necessarily be artificial. Therefore, we will have to assume the existence of the positive definite limiting matrix, which is a modest assumption.

4 Real data application

We have used the same data as in Elfving (2000), where the model (2) was used. Taking into the account the large amount of data, it was of interest for us to calculate the MLE of the parameters of the new model and to compare the results with those obtained from the simpler model.

The data used by Elfving consisted of 2333 remeasured trees on permanent plots in the Swedish National Forest Inventory. Together with the diameter,

numerous additional variables describing both the stand and the site were registered. Among them were the basal area of larger trees, field vegetation type, soil depth, texture, moisture, etc. Since our objective was to improve the simpler model - the one without measurement errors, we included the same set of explanatory variables as in the original paper. The number of the parameters in the vector $\boldsymbol{\beta}$ is 13, thus bringing the total number of parameters to be estimated to 15. The list of the parameters together with the descriptions of the corresponding explanatory variables can be seen in Table 1. The function $g(d)$ relating the measurement error for growth in basal area to the diameter (see (4)), was taken from the paper by Elfving. There the relation $g(d) = 0.026(d+5)^2$ was obtained. The quadratic shape of this relation is due to the known linear shape of the relation between the measurement error for the diameter and the diameter. The latter was estimated using the repeated independent measurements made on a subset of plots within a week.

4.1 Random optimization

The density function of the LNN does not take form of any standard well-known integral. Therefore we opted for a numerical approach to the estimation of the parameters of the regression model (3). The likelihood function of the sample was maximized with respect to $\boldsymbol{\beta}, \sigma, \tau$ by means of *random optimization* (see, for example, Lee and Rhinehart (1998)). This method operates as follows. The value of the likelihood is calculated in a reasonably chosen starting point $P_0 = (\boldsymbol{\beta_0}, \sigma_0, \tau_0)$. This starting point may obtained from the simpler model such as (2). Then a random step in the parameter space is made, producing another point $P_1 = (\boldsymbol{\beta_1}, \sigma_1, \tau_1)$. The value of the likelihood is calculated in P_1 and compared to that in P_0. If it is larger, P_1 is established as a current MLE of the parameters. The algorithm continues until some sort of stability is obtained. Although, there is no guarantee that the obtained maximum is global, this and similar approaches are often used in applications. The random optimization is rather demanding in terms of computer resources. Our computations required approximately 300 hours of CPU time on a Pentium II PC.

In order for the random optimization procedure to converge in finite time, it is necessary to impose some inequality-type restrictions (e.g. bounds) on the parameters. Since it was the practical application that motivated our study, we were able to do that by utilizing the prior knowledge about the values of the parameters.

As a starting point for the parameters of regression we utilized the estimates derived from model (2) by means of the weighted least squares.

Apparently, there is certain vagueness in the choice of the random steps. The first random steps in the parameter space were chosen to be one-dimensional Gaussian for each parameter. The variance was chosen to be approximately 30% of the current value of the parameter estimate. After several thousands Gaussian steps the value of the MLE stabilized. One could suppose that the

obtained MLE was close to the real one. Then the uniform distribution was employed to search in the m-dimensional cubic neighborhood of the current MLE. The range of the uniform distribution was gradually narrowed, until the the required accuracy was achieved. Note that since we based our inference on the imprecise sample data, it was pointless to strive for an extremely high accuracy of the estimates. In fact, adding or removing a small fraction of observations changed the estimates already in third digit, making it unreasonable to try to achieve higher precision. Moreover, taking into account the accuracy of the numerical evaluation of the integrals in the likelihood function, we were able to obtain three accurate digits in the estimates.

4.2 Results and conclusion

Both the initial values of the parameters (those obtained in Elfving (2000) from model (2)) and the obtained values of the MLE can be found in Table 1. We note that none of the coefficients of the regression function have changed the sign. This fact is encouraging, since the opposite would mean that the effects of the corresponding variables are reversed, which should not happen. As expected, the model supports our prior knowledge of the role of predictors. Since, as we have mentioned in Section 2, model (2) is somewhat incongruous with the data, it was of most interest for us to compare our model (3) with the original model (1). The answer to the question whether the new model helps to explain the data better, is affirmative for the following reason. The new model *incorporates* the original model (1), since it extends the model (1) by taking into consideration the measurement errors (i.e. accounting for cases with negative recorded growth). Clearly, if the measurement errors were small (e.g. $\tau \approx 0$), (3) would transform into (1). Therefore, we can compare the two models directly by means of the likelihood ratio test. To do so, we have to restrict ourselves to the subset of data with *positive* recorded growth, e.g. where model (1) is applicable. The likelihood ratio test statistic $\chi^2 = -2 \log \frac{\sup_\theta L_n^0}{\sup_\theta L_n^1}$, where L_n^0 and L_n^1 are the likelihood functions of models (1) and (3), respectively, is approximately Chi-square distributed, should model (1) be correct. We obtained $\chi^2 \approx -2 \log 0.0001 \approx 18.42$, which is larger than 99.99% percentile of Chi-square distribution with one degree of freedom. This proves the necessity to include measurement errors in forest growth models.

Acknowledgements

The authors would like to thank Professor B. Elfving for introducing the problem, and Professor Y. Belyaev, Dr J. Yu and Dr M. Ekström for helpful comments. We would like to thank the Swedish Foundation for Strategic Environmental Research, MISTRA, for financial support. This research was conducted using the resources of High Performance Computing Center North (HPC2N).

References

Candy, S. (1997). Estimation in forest yield models using composite link functions with random effects. *Biometrics* **53**, 146–160.

Ek, A. R. and Monserud, R. A. (1974). FOREST: a computer model for simulating the growth and reproduction of mixed-species forest stands. Technical Report R 2635, Coll. of Agr. and Life Sci., School of Natural Resources, University of Wisconsin.

Elfving, B. (2000). Individual-tree basal area growth functions based on data from remeasured inventory plots in Sweden. Manuscript.

Hawkins, D. (1991). The convolution of the normal and lognormal distributions. *South African Statistical Journal* **25**, 99–128.

Hoadley, B. (1971). Asymptotic properties of ML estimators for independent not identically distributed case. *The Annals of Mathematical Statistics* **42**, 1977–1991.

Lee, J. and Rhinehart, R. (1998). Heuristic random optimization. *Computers chem. Engng* **22**, 427–444.

Roussas, G. (1968). Asymptotic normality of the maximum likelihood estimate in markov processes. *Metrika* **14**, 62–70.

Siitoen, M. (1995). The MELA system as a forestry modelling framework. *Lesnictvi-Forestry* **41**, 173–178.

Stage, A. R. (1974). Prognosis model for stand development. Technical Report INT-137, USDA For.Serv.Res.Pap.

Wald, A. (1949). Note on the consistency of the ML estimate. *The Annals of Mathematical Statistics* **20**, 595–601.

Appendix

Derivatives of LNN density function

The first and the second derivatives of the LNN density function in forest growth model with measurement errors taken with respect to the parameters take the following form ($A_k(t)$ is the same as in section "Model"):

$$
\begin{aligned}
\frac{\partial f_k}{\partial \beta_j} &= \int_{-\infty}^{\infty} \frac{x_{kj}}{\sigma^2} \left(t - \sum_j \beta_j x_{kj} + \frac{\sigma^2}{2} \right) A_k(t) dt \\
\frac{\partial f_k}{\partial \sigma^2} &= \int_{-\infty}^{\infty} \frac{1}{\sigma^4} \left((t - \sum_j \beta_j x_{kj})^2 - \frac{\sigma^4}{4} - \sigma^2 \right) A_k(t) dt \\
\frac{\partial f_k}{\partial \tau^2} &= \int_{-\infty}^{\infty} \frac{1}{2g(\mathbf{x}_k)\tau^4} \left((e^t - Y_k)^2 - g(\mathbf{x}_k)\tau^2 \right) A_k(t) dt \\
\frac{\partial^2 f_k}{\partial \beta_j \partial \beta_l} &= \int_{-\infty}^{\infty} \frac{x_{kj} x_{kl}}{\sigma^4} \left((t - \sum_j \beta_j x_{kj} + \frac{\sigma^2}{2})^2 - \frac{x_{kj} x_{kl}}{\sigma^2} \right) A_k(t) dt
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 f_k}{\partial\beta_j\partial\sigma^2} &= \int_{-\infty}^{\infty}\frac{x_{kj}}{\sigma^4}\left(-(t-\sum_j\beta_j x_{kj})-\frac{1}{2}(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})+\right.\\
&\left.+\frac{1}{2\sigma^2}(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})^2(t-\sum_j\beta_j x_{kj}-\frac{\sigma^2}{2})\right)A_k(t)dt\\
\frac{\partial^2 f_k}{\partial(\sigma^2)^2} &= \int_{-\infty}^{\infty}\frac{1}{4\sigma^6}\left(3\sigma^2-3(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})(t-\sum_j\beta_j x_{kj}-\frac{\sigma^2}{2})+\right.\\
&+\frac{1}{\sigma^2}(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})^2(t-\sum_j\beta_j x_{kj}-\frac{\sigma^2}{2})^2-\\
&\left.-\sigma^4-3(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})^2+3\sigma^2(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})\right)A_k(t)dt\\
\frac{\partial^2 f_k}{\partial\beta_j\partial\tau^2} &= \int_{-\infty}^{\infty}\frac{x_{kj}}{2\sigma^2 g(\mathbf{x}_k)\tau^4}((e^t-Y_k)^2-g(\mathbf{x}_k)\tau^2)(t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})A_k(t)dt\\
\frac{\partial^2 f_k}{\partial\sigma^2\partial\tau^2} &= \int_{-\infty}^{\infty}\frac{1}{4\sigma^4 g(\mathbf{x}_k)\tau^4}((e^t-Y_k)^2-g(\mathbf{x}_k)\tau^2)\times\\
&\times((t-\sum_j\beta_j x_{kj}+\frac{\sigma^2}{2})(t-\sum_j\beta_j x_{kj}-\frac{\sigma^2}{2})-\sigma^2)A_k(t)dt\\
\frac{\partial^2 f_k}{\partial(\tau^2)^2} &= \int_{-\infty}^{\infty}\frac{1}{4g^2(\mathbf{x}_k)\tau^8}\left((e^t-Y_k)^4-6g(\mathbf{x}_k)\tau^2(e^t-Y_k)^2+3g^2(\mathbf{x}_k)\tau^4\right)A_k(t)dt
\end{aligned}
$$