

The partly parametric and partly nonparametric additive risk model



Nils Lid Hjort (joint work with Emil Aas Stoltenberg)
published in *Lifetime Data Analysis*, Nov. 2021

Department of Mathematics, University of Oslo

Ørnulf Borgan $2^3 \times 3^2$ days, Univ of Oslo, 13-14/vi/2022

Supplementary details to Sven Ove's familietre

Known Ørnulf and Odd from ca. 1973. Since ca. 1980, learned counting processes, martingales, survival analysis from them (seminars, reading, conversations, blackboarding).

Invited discussant to Andersen and Borgan big SJS paper, Bolkesjø, 1984 – where I wrote five papers in the process, each with a bit of Ørnulf and Odd in them. In particular, Nils SJS 1986 is a ... natural follow-up to Ørnulf SJS 1984 (I generalise from 'at the model' to 'outside the model', add on covariates, Bayes, etc.).

Looking back ...

(i) The 'Oslo survival and event history group' grew a strong one, in the 1980ies, 1990ies, to the present. Interconnected branching processes of lots o' scholars. From my own circles: Odd, Ørnulf, Nils, Inge H, Sven Ove, Håkon G, many of their students again; 'my own' Martin Jullum, Céline Cunen, Kristoffer Hellton, Jens Kristoffer Haug, Emil Stoltenberg, Catharina Stoltenberg, Marthe Aastveit.

(ii) The Ørnulf–Odd terrain has become one of my own 4-5 favourite research directions.

Cases in point:

Nils, Annals of Statistics, 1990, goodness-of-fit, pages 1221–1258:

Acknowledgements. Conversations with Ørnulf Borgan and Odd Aalen triggered some of the research that led to the present paper. I am grateful for the careful work of an associate editor and the encouraging persistence of editor van Zwet, which finally caused me to carry out the necessary revision.

Nils, Annals of Statistics, 1990, Beta processes, pages 1259–1294:

Acknowledgements. Conversations with Odd Aalen and Ørnulf Borgan triggered some of the research that led to the present paper. I am grateful for the careful work of an associate editor and the persistent encouragement of editor van Zwet, which finally caused me to carry out the necessary revision.

D.R. Multiplicative Cox and O.O. Linear Aalen

Survival data $(t_i, z_{i,1}, \dots, z_{i,r}, \delta_i)$, modelling hazard rates

$h_i(s) = h(s | z_i)$: Cox says

$$h_i(s) = h_0(s) \exp(z_i^t \beta) = h_0(s) \exp(z_{i,1} \beta_1 + \dots + z_{i,r} \beta_r),$$

Aalen says

$$h_i(s) = z_i^t \alpha(s) = z_{i,1} \alpha_1(s) + \dots + z_{i,r} \alpha_r(s).$$

Aalen methods give full nonparametric estimation and inference for cumulative hazard component functions $A_j(t) = \int_0^t \alpha_j(s) ds$.

Full nonparametric control:

$$H_i(t) = z_i^t A(t) = z_{i,1} A_1(t) + \dots + z_{i,r} A_r(t),$$

$$S(t | z) = \exp\{-z_1 A_1(t) - \dots - z_r A_r(t)\},$$

$$\tilde{S}(t | z) = \exp\{-z_1 \tilde{A}_1(t) - \dots - z_r \tilde{A}_r(t)\}.$$

Hjort and Stoltenberg (LIDA, Nov 2021):

some parametric: $\alpha_1(s, \theta_1), \dots, \alpha_p(s, \theta_p),$

some nonparametric: $\alpha_{p+1}(s), \dots, \alpha_{p+q}(s).$

Core ideas

The **classic Aalen model** (1980, 1989, 1993, ABGK 1993, ABG 2008, Martinussen and Scheike 2007):

$$h_i(s) = z_{i,1}\alpha_1(s) + \cdots + z_{i,r}\alpha_r(s),$$

where the $\alpha_1(s), \dots, \alpha_r(s)$ are taken **nonparametric**.

There are estimators $\tilde{A}_j(t)$ for $A_j(t) = \int_0^t \alpha_j(s) ds$:
one has **Gaussian process limits**

$$\sqrt{n}\{\tilde{A}_j(t) - A_j(t)\} \rightarrow_d L_j(t),$$

etc. But **estimators become erratic** for growing t ,
particularly if r is not small.

Can also do **fully parametric**, $\alpha_j(s, \theta_j)$, for $j = 1, \dots, r$,

$$\ell_n(\theta) = \sum_{i=1}^n \int_0^\tau \{\log h_i(s, \theta) dN_i(s) - Y_i(s)h_i(s, \theta) ds\}$$

becomes the log-likelihood in $\theta = (\theta_1, \dots, \theta_r)$. But this is often
too constrained or **too complicated**.

We're aiming for partly para partly nonpara

We go for **some para**, **some nonpara**:

$\alpha_1(s, \theta_1), \dots, \alpha_p(s, \theta_p), \alpha_{p+1}(s), \dots, \alpha_{p+q}(s)$.

If we **succeed** (and we do ...):

- ♠ better precision and interpretation for the **para parts**;
- ♠ better precision also for **nonpara parts**;
- ♠ better precision for 'what matters', the survival curves

$$\begin{aligned}\widehat{S}(t | z) = \exp\{ & -z_1 A_1(t, \widehat{\theta}_1) - \dots - z_p A_p(t, \widehat{\theta}_p) \\ & - z_{p+1} \widehat{A}_{p+1}(t) - \dots - z_{p+q} \widehat{A}_{p+q}(t) \} ;\end{aligned}$$

- ♠ the start of a scheme for model construction and model selection;
- ♠ better survival analysis.

What we have to do

- ♠ inventing estimation procedures
(first for **para**, then for **nonpara** given these);
- ♠ analysing and understanding their behaviour;
- ♠ sorting out **complicated asymptotics**;
- ♠ estimating **variances and covariances**, etc.;
- ♠ checking performance, how much have we won, etc:
- ♠ devising **goodness-of-fit tests** for the parametric parts;
- ♠ implementing and demonstrating how well things work.

This is partly **a tall order**.

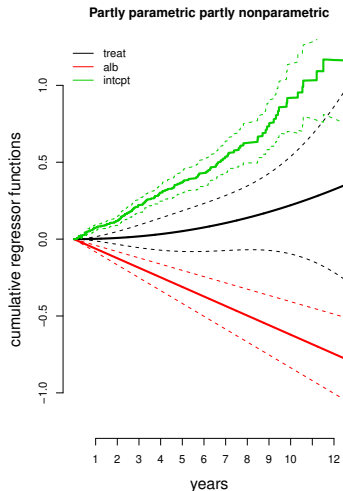
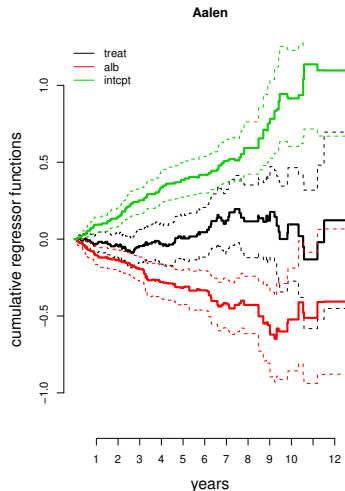
Growing literature on extensions and variations on the Aalen linear risk model, but essentially **still nonparametric**, and not much in these Nils-Emil directions.

McKeague and Sasieni (1994): **some α_j constant**, the rest nonparametric – but their estimation methods don't generalise easily, so we've started afresh.

PBC patients study

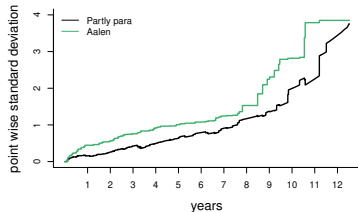
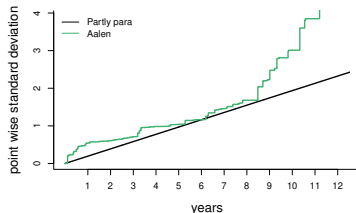
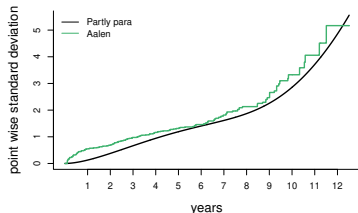
312 patients, Mayo Clinic, treatment/placebo, albumin level:

$$h_i(s) = abs^{b-1} \text{treat}_i + \theta_a \text{alb}_i + \alpha_3(s)$$



We've lowered the variances

Pointwise standard deviations, Aalen vs. Nils-Emil, when estimating $A_1(t) = at^b$, $A_2(t) = \theta_a t$, $A_3(t)$:



Fully nonparametric scheme

Classic Aalen: **fully nonparametric**,

$$\tilde{A}(t) = \int_0^t G_n(s)^{-1} n^{-1} \sum_{i=1}^n w_i(s) z_i \, dN_i(s),$$

with $G_n(s) = n^{-1} \sum_{i=1}^n Y_i(s) w_i(s) z_i z_i^t$. Reasonably standard notation: z_i covariate vector for no. i ,

$$Y_i(s) = I(\text{no. } i \text{ is at risk at time } s),$$

$$dN_i(s) = I(\text{no. } i \text{ is observed to jump in } [s, s + ds]),$$

$w_i(s)$ is a previsible weight function; good default: $w_i(s) = 1$.

We block these into $\tilde{A}_{(1)}(t)$, the p first, and $\tilde{A}_{(2)}(t)$, the q last.

We estimate the para parts $\alpha_1(s, \theta_1), \dots, \alpha_p(s, \theta_p)$

by closeness to $\tilde{A}_1, \dots, \tilde{A}_p$.

Estimating the para components

Different ideas might work ... our favoured class is this:

Aiming at minimising

$$\int_0^\tau \{\alpha_{(1)}(s, \theta) - \alpha_{(1)}(s)\}^t V_n(s) \{\alpha_{(1)}(s, \theta) - \alpha_{(1)}(s)\} ds,$$

with suitable $p \times p$ matrix $V_n(s)$. Multiplying out parts with θ , and estimating: we define $\hat{\theta}$ to be the minimiser of

$$\int_0^\tau \alpha_{(1)}(s, \theta)^t V_n(s) \alpha_{(1)}(s, \theta) ds - 2 \int_0^\tau \alpha_{(1)}(s, \theta)^t V_n(s) d\tilde{A}_{(1)}(s).$$

Our default choice is

$$V_n(s) = n^{-1} \sum_{i=1}^n Y_i(s) z_{i,(1)} z_{i,(1)}^t.$$

Backfitting for estimating the nonpara components

We exploit the $\alpha_1(s, \hat{\theta}_1), \dots, \alpha_p(s, \hat{\theta}_p)$ to estimate $A_{p+1}, \dots, A_{p+q}$ better:

$$\begin{aligned} d\hat{A}_{(2)}(s) = & G_{n,22}(s)^{-1} n^{-1} \sum_{i=1}^n w_i(s) z_{i,(2)} \\ & \{dN_i(s) - Y_i(s) z_{i,(1)}^t \alpha_{(1)}(s, \hat{\theta}) ds\}. \end{aligned}$$

This yields $\hat{A}_{p+1}(t), \dots, \hat{A}_{p+q}(t)$ – better than nonpara Aalen $\tilde{A}_{p+1}(t), \dots, \tilde{A}_{p+q}(t)$ (if the para models are ok).

Large-sample results for the para part

Somewhat tough going, but we manage, via tools of the trade – Taylor expansion; remainder terms; martingales; behaviour of nonpara $d\tilde{A}(s)$.

After deciding on a $p \times p$ function $V_n(s)$, $\hat{\theta}$ is the minimiser of

$$\int_0^T \alpha_{(1)}(s, \theta)^t V_n(s) \alpha_{(1)}(s, \theta) ds - 2 \int_0^T \alpha_{(1)}(s, \theta)^t V_n(s) d\tilde{A}_{(1)}(s),$$

and the solution to $S_n(\hat{\theta}) = 0$, with

$$S_n(\theta) = \int_0^T \alpha_{(1)}^*(s, \theta)^t V_n(s) \{d\tilde{A}_{(1)}(s) - \alpha_{(1)}(s, \theta) ds\}.$$

Result I: under para model conditions,

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \Gamma^{-1} S \sim N_m(0, \Gamma^{-1} \Omega \Gamma^{-1}),$$

with clear descriptions of Γ and Ω .

More on the para estimation part

The **Result 1** works for any choice of $p \times p$ weight function $V_n(s)$. There is an optimal choice, for (very) large n :

(i) $\Gamma^{-1}\Omega\Gamma^{-1}$ is as small as possible with

$$V(s) = Q(s)^{-1}, \quad \text{with} \quad Q(s) ds = [G(s)^{-1} dH(s) G(s)^{-1}]_{11},$$

with G and H associated with the nonpara Aalen estimator. This leads to $\Gamma = \Omega = \Omega_0$, and limit variance matrix

$$\Omega_0^{-1} = \left\{ \int_0^\tau \alpha_{(1)}^*(s, \theta_0)^t Q(s)^{-1} \alpha_{(1)}^*(s, \theta_0) ds \right\}^{-1}.$$

(ii) This Ω_0^{-1} is as small as possible with $Q(s) ds$ being as small as possible, which can be handled by clever $G(s), H(s)$, involving also nonpara $\alpha(s)$ estimation.

(iii) This is the best possible limit distribution (i.e. over all estimators for θ , also outside our $V_n(s)$ weight function class).

Large-sample results for the nonpara part

Again somewhat tough going, but we manage, tools of the trade.
Need

$$\phi_n(s) = n^{-1} \sum_{i=1}^n Y_i(s) w_i(s) z_{i,(2)} z_{i,(1)}^t \alpha_{(1)}^*(s, \theta_0) \rightarrow_{\text{pr}} \phi(s).$$

Result II: with $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d \Gamma^{-1}S$, a few Gaussian processes, etc.,

$$\begin{aligned} \sqrt{n}\{\hat{A}_{(2)}(t) - A_{(2)}(t)\} &\rightarrow_d \int_0^t G_{22}(s)^{-1} dU_{(2)}(s) \\ &\quad - \int_0^t G_{22}(s)^{-1} \phi(s) ds \Gamma^{-1}S. \end{aligned}$$

May now bind together **Result I** and **Result II**, with delta method tools, to find limiting normal distribution of any

$$\hat{\mu} = \mu(A_1(\cdot, \hat{\theta}_1), \dots, A_p(\cdot, \hat{\theta}_p), \hat{A}_{p+1}(\cdot), \dots, \hat{A}_{p+q}(\cdot)).$$

We're in business.

Goodness-of-fit monitoring processes

For suitable weight functions $K_{n,j}(s)$, we work with

$$R_{n,j}(t) = \sqrt{n} \int_0^t K_{n,j}(s) \{d\tilde{A}_j(s) - \alpha_j(s, \hat{\theta}) ds\},$$

monitoring difference between nonpara and para.

Result III: Under para model conditions, there are well-defined Gaussian zero-mean process limits, $R_{n,j}(t) \rightarrow_d R_j(t)$.

This leads to **goodness-of-fit tests** of different types (chi-squared, or Kolmogorov–Smirnov, etc.): can test the hypothesis that $\alpha_j(s, \theta_j)$ is ok.

The rest of the Nils-Emil LIDA paper

- ♠ Estimators for all quantities in limit distributions.
- ♠ Looking into relative efficiencies (how much have we won?).
- ♠ Implementation, simulations (the theory works!), application.
- ♠ More on the monitoring goodness-of-fit processes.

Also, the start of a **FIC machinery**: with μ a focus parameter (function of 'everything' put into the model), we have many different estimates $\hat{\mu}_{\text{choice}}$, depending on **choice**: **which** A_j are para (and which para models), **which** A_j are nonpara.

The **FIC** is estimating

$$\text{risk}_{\text{choice}} = E_{\text{true}} (\hat{\mu}_{\text{choice}} - \mu_{\text{true}})^2$$

for each of the candidate choices.

Remarks

- ♣ Can have time-dependent covariates $z_i(s)$ (but be careful with interpretation of $h_i(s)$).
- ♣ Parallel to GAM: Hastie, Tibshirani, others, big success story:

$$E(Y | x_1, \dots, x_r) = f_1(x_1) + \dots + f_r(x_r),$$

with nonpara fitting. We could use some of our ideas for having **some para some nonpara** in GAM.

- ♣ More to do on the estimation side (though we do have asymptotic optimality results for versions of our schemes).
- ♣ More to do on the **goodness-of-fit** arena.
- ♣ For a **full-fledged R package (?)**, would need model selection. AIC, BIC don't work – but **can build FIC things** (but not easy).
- ♣ A **para-nonpara cure model**.
- ♣ **Glattting av glidende gjennomsnitt**: can do nonpara for $\hat{\alpha}_{p+1}(s), \dots, \hat{\alpha}_{p+q}(s)$.

Semiparametric estimation of (a single) cumulative hazard

Above: covariates $1, \dots, r$, some A_j fully para, some A_j fully nonpara. Different type of para-nonpara idea: for classical $(t_1, \delta_1), \dots, (t_n, \delta_n)$ survival setup, use

$$\hat{A}(t) = \int_0^t \left[\{1 - w_n(s)\} \frac{dN(s)}{Y(s)} + w_n(s) \alpha(s, \hat{\theta}) ds \right],$$

with $w_n(s) \in [0, 1]$ a weight function.

We wish $w_n(s) \rightarrow 1$ for growing s .

Details – motivation, theory, choice of w_n , applications – in Nils and Emil (2022, new paper). These ideas may also be used in Nils and Emil (LIDA 2021, the present talk).

A few references

- ♠ ABGK (1993). *Yellow Brick Road*.
- ♠ Hjort, N.L., Stoltenberg, E.Aa. (2021). The partly parametric and partly nonparametric additive risk model. *Lifetime Data Analysis*.
- ♠ Hjort, N.L., Stoltenberg, E.Aa. (2022). Semiparametric estimation of hazard rates. *Manuscript*.
- ♠ Martinussen, T., Scheike, T. H. (2007). *Dynamic Regression Models for Survival Data*. Springer Verlag.
- ♠ McKeague, I.W., Sasieni, P.D. (1994). A partly parametric additive risk model. *Biometrika*.
- ♠ Stoltenberg, E.A. (2020). The standard cure model with a linear hazard. *arXiv*
- ♠ Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics*.
- ♠ Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*.
- ♠ Aalen, O.O. (1993). Further results on the nonparametric linear regression model in survival analysis. *Statistics in Medicine*.
- ♠ Aalen, O.O., Borgan, Ø., Gjessing, H. (2008). *Survival and Event History Analysis*. Springer Verlag.