

Modelling fertility trait as censored observations in Finnish Ayrshire data

K. Matilainen, G. P. Aamand, E. A. Mäntysaari

Background

- Incomplete data may cause bias for fertility evaluation of young sires:
 - Best fertile daughters have records available first.
 - Observations of poor fertile daughters come late (or not at all).
- ➔ Inclusion of censored observations to the data. Possible right censored observations in Nordic fertility evaluation:
 - Interval from calving to first insemination (ICF): Calving is known, but first insemination not yet occurred.
 - Interval from first to last insemination (IFL): One or several inseminations, but uncertain pregnancy status after last recorded insemination.

Current evaluation

- Censored IFL records are extended. Extension is done in pre-processing, and the penalty
 - Is based on breed, country and lactation wise averages
 - Is dependent on occurrence of pregnancy test and time from the last recorded insemination
- ➔ Another option would be to use right censored Gaussian model.

Right censored Gaussian model

- During the analysis, censored records are replaced by generated records w using truncated normal distribution with the information on current solutions for model effects and observed censored values c :

$$w_i \mid b, u, \sigma_u^2, \sigma_e^2, c_i \sim \mathcal{N}(x_i' b + z_i' u, \sigma_e^2) I(w_i \geq c_i)$$

- Leads to two step procedure where generated observations are updated from time to time based on new solutions for model effects.
- Contemporary groups must be constructed carefully to avoid situations that all observations in the same group are censored.

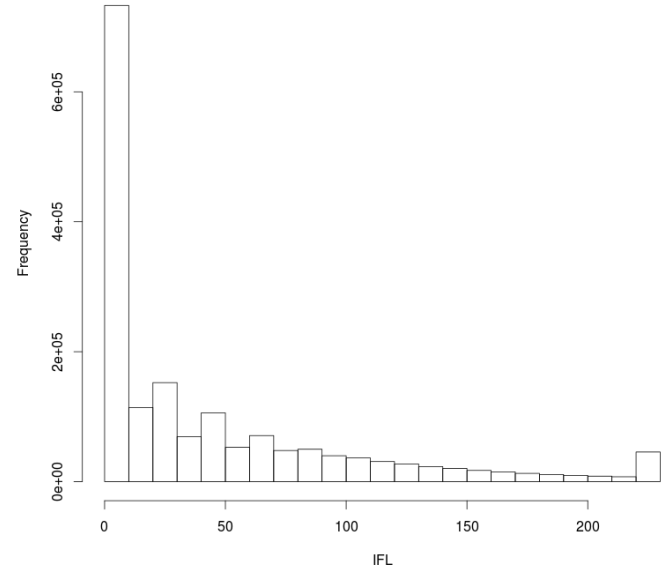
Aim of this study

- To investigate three approaches for first parity IFL records:
 1. Linear model with original non-censored and censored observations
 2. Linear model with non-censored observations and extended values for censored observations
 3. Right censored Gaussian model



Field data

- Data contained 1.7 million IFL records for Finnish Red Dairy Cattle first parity cows from 1992 to 2018.
- 14% of the observations censored, i.e., unknown pregnancy status or without subsequent calving information.
 - For censored model analysis, 0.6% of the observations were removed to prevent problem with fixed effect classes that have all observations censored.
- Pedigree contained 3 million animals.



Analyses

- All studied analyses
 - had the same model effects as in the Nordic evaluations
 - had the same variance parameters with heritability 0.05
 - were made by modified MiX99, which uses iterative preconditioned conjugate gradient algorithm (PCG) to solve mixed model equations



Convergence

- Example of computing costs of different analyses to convergence (relative difference $< 10^{-6}$ and one PCG iteration within a round for censored model)

	PCG iterations	Time (min)
Original	699	5
Extended	661	5
Censored	2225 in 31 rounds	27

Breeding values

- For 2869 Nordic RDC sires:
 - Correlations (cor) of estimated breeding values (EBV) between analyses were high.
 - Proportion (prop) of the same sires among the best 5% showed some differences.
 - Genetic trend (standardized and reversed, i.e., higher the better) turned positive in 2008!

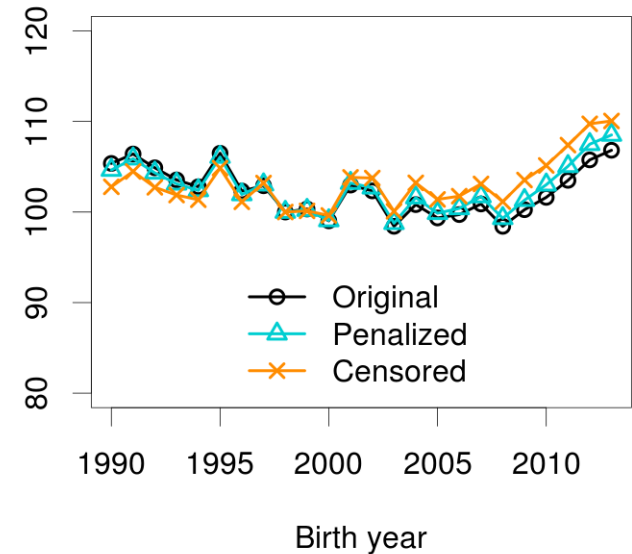
	cor	prop (%)
Original vs. Extended	0.99	90
Original vs. Censored	0.95	75
Extended vs. Censored	0.97	81

Breeding values

- For 2869 Nordic RDC sires:
 - Correlations (cor) of estimated breeding values (EBV) between analyses were high.
 - Proportion (prop) of the same sires among the best 5% showed some differences.
 - Genetic trend (standardized and reversed, i.e., higher the better) turned positive in 2008!

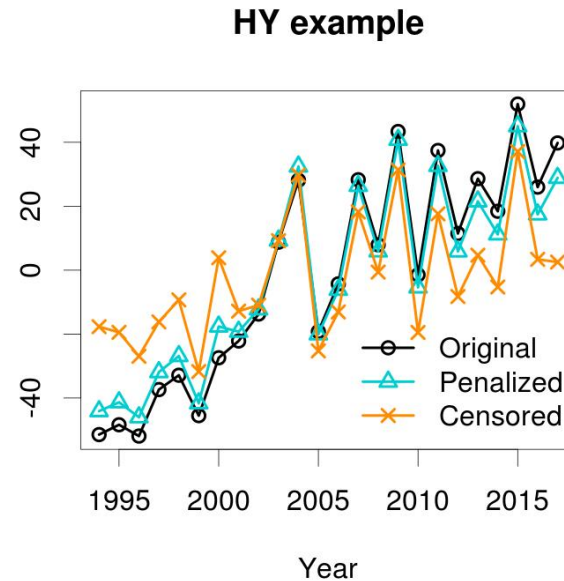
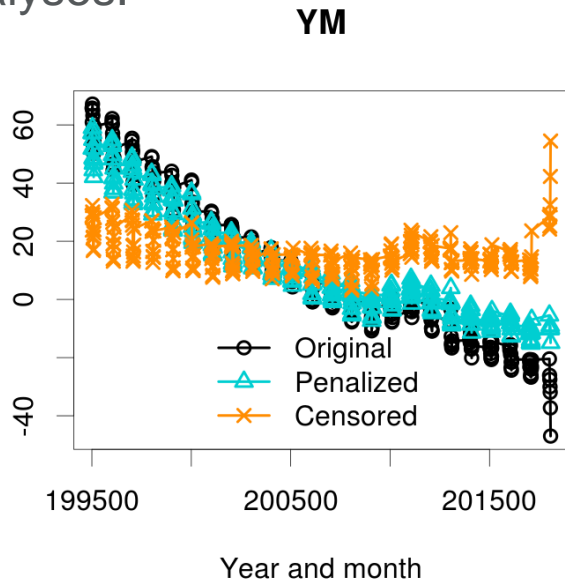
	cor	prop (%)
Original vs. Extended	0.99	90
Original vs. Censored	0.95	75
Extended vs. Censored	0.97	81

Genetic trend



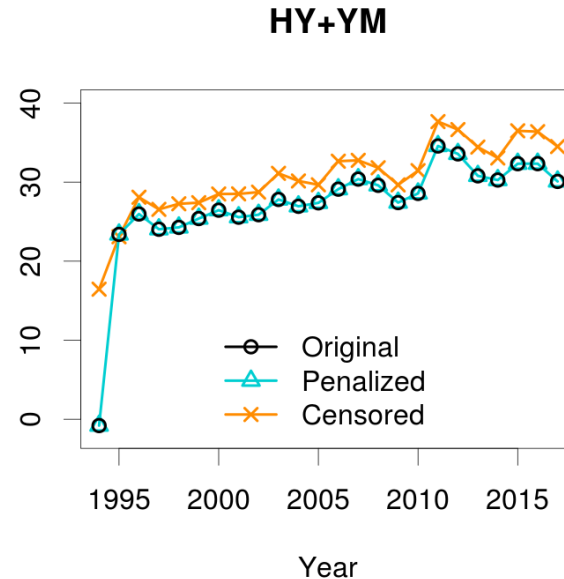
Fixed effects related to year

- Year \times Month and Herd \times Year solutions had different patterns between analyses.



Fixed effects related to year

- Year \times Month and Herd \times Year solutions had different patterns between analyses.
- But average yearly solutions based on these effects had the same pattern.



Validation

- Deregressed genetic evaluations (DRP) were regressed on EBV from reduced data with 4 year cut off.
 - 362 validation bulls having effective record contribution $ERC=0$ based on reduced data and $ERC>10$ based on full data.

	R^2	b_1
Original	0.08	0.75
Extended	0.09	0.83
Censored	0.12	0.93

$$DRP_{2018} = b_0 + b_1 EBV_{2014}$$

$$R^2 = R_m^2 / w$$

$$w = \text{mean DRP accuracy} = 0.57$$

- Note: Simple data cut didn't have the possibility to indicate "true" censoring of observations four years back.

Examination with 2016 data

- Correlations of EBVs for 34 bulls that have at least 10 daughters without censored observations in 2018, but some daughters with censored observations in 2016 data (proportion of censored observations was 0.40).
 - A. 21 bulls from 34, which have > 5 daughters in 2016 (0.24).
 - B. 13 bulls from 34, which have ≤ 5 daughters in 2016 (0.66).

	All	A.	B.
Original	0.80	0.87	0.68
Extended	0.81	0.86	0.71
Censored	0.84	0.87	0.81

Examination with 2016 data

- Correlations of EBVs for 34 bulls that have at least 10 daughters without censored observations in 2018, but some daughters with censored observations in 2016 data (proportion of censored observations was 0.40).
 - A. 21 bulls from 34, which have > 5 daughters in 2016 (0.24).
 - B. 13 bulls from 34, which have ≤ 5 daughters in 2016 (0.66).

	All	A.	B.
Original	0.80	0.87	0.68
Extended	0.81	0.86	0.71
Censored	0.84	0.87	0.81

Conclusions



- For univariate IFL analysis
 - + Right censored Gaussian model was feasible.
 - + Censored model corrected estimates of fixed effects and breeding values especially for sires with large proportion of censored observations.
 - + Also model validation showed that censored model improved the evaluation.
 - Censored model is always more time consuming compared to linear models.

Acknowledges

Data:

NAV Nordic Cattle Genetic Evaluation



Funding:

MMM, NAV, Viking Genetics, Faba and Valio



Thank you!