Can Machine Learning Support Survival Model Selection to Inform Economic Evaluations?

MSR17

Exploring k-Fold Cross Validation Based Model Selection in Seven Datasets

Inigo Bermejo[1] and Sabine Grimm[2]

[1]Data Science Institute, Hasselt University, Belgium; [2]Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, The Netherlands.

Background & Aim

- Selection of survival models for informing economic evaluations with limited long-term data relies on metrics of statistical goodness of fit (AIC/BIC).
- Issue: models selected based on full trial data might underperform in the target population due to overfitting.
- k-fold cross validation (CV), commonly used in machine learning, allows for estimation of fit in unseen data.
- We explore whether k-fold CV improves model selection.

Methods

- Data: 7 public long-term survival datasets covering a range of diseases.
- Preprocessing: We simulated 100 artificial data locks by sampling 250 patients without replacement, and right-censoring once median survival was reached.
- Experiments:
  - Fit parametric and flexible survival models to each simulated dataset
  - Compare models with lowest AIC/BIC as estimated using traditional methods vs 10-fold CV.
- Evaluation: restricted mean survival time (RMST) error of best-fitting models relative to the RMST calculated from the full dataset's Kaplan-Meier (KM).

Survival models
Loglogistic
Weibull
Lognormal
Gamma
Exponential
Gompertz
Generalised gamma
Spline k=1-4 hazard
Spline k=1-4 odds
Spline k=1-4 normal

Results

Figure 1: KM of the seven datasets show heterogeneity

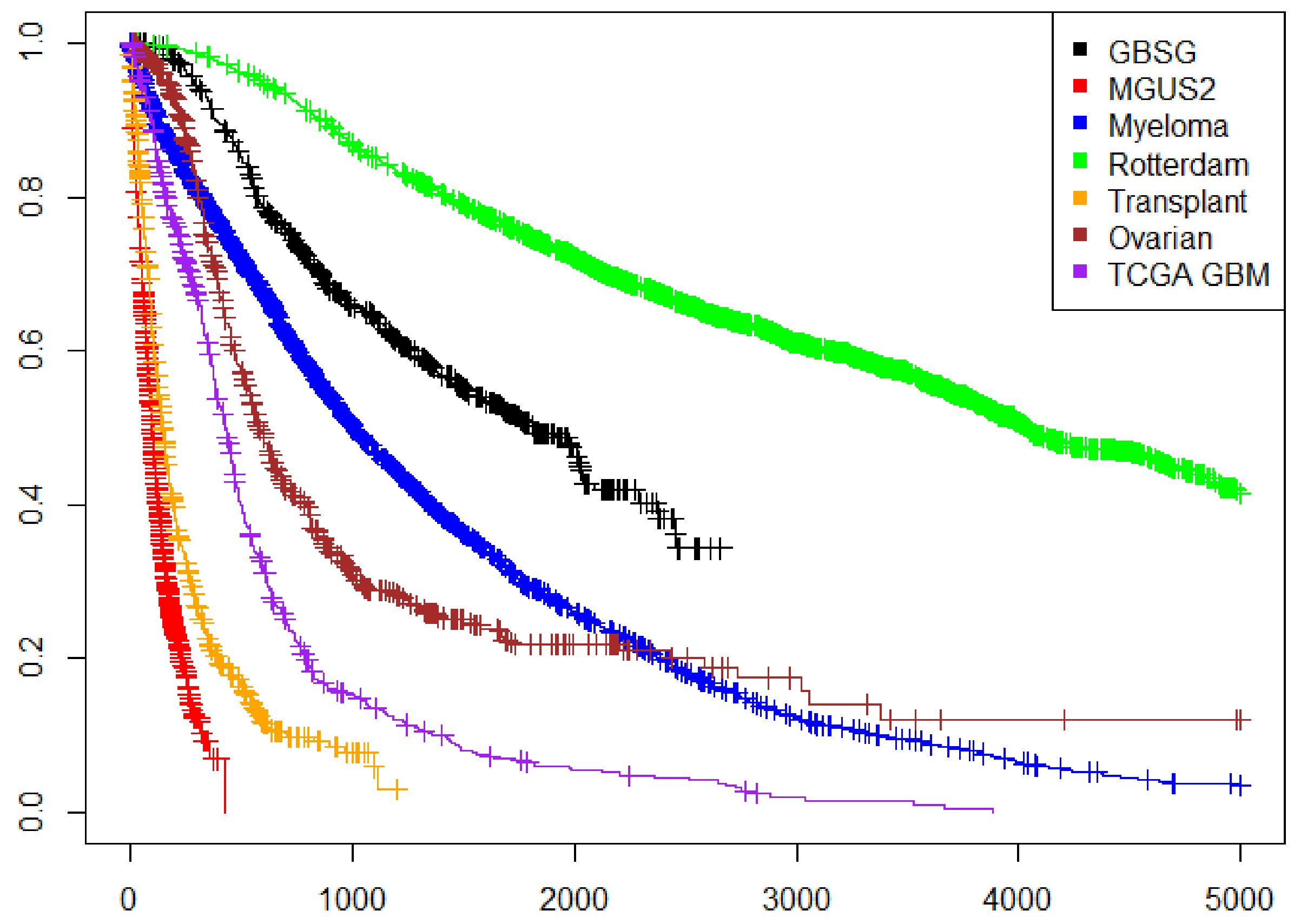
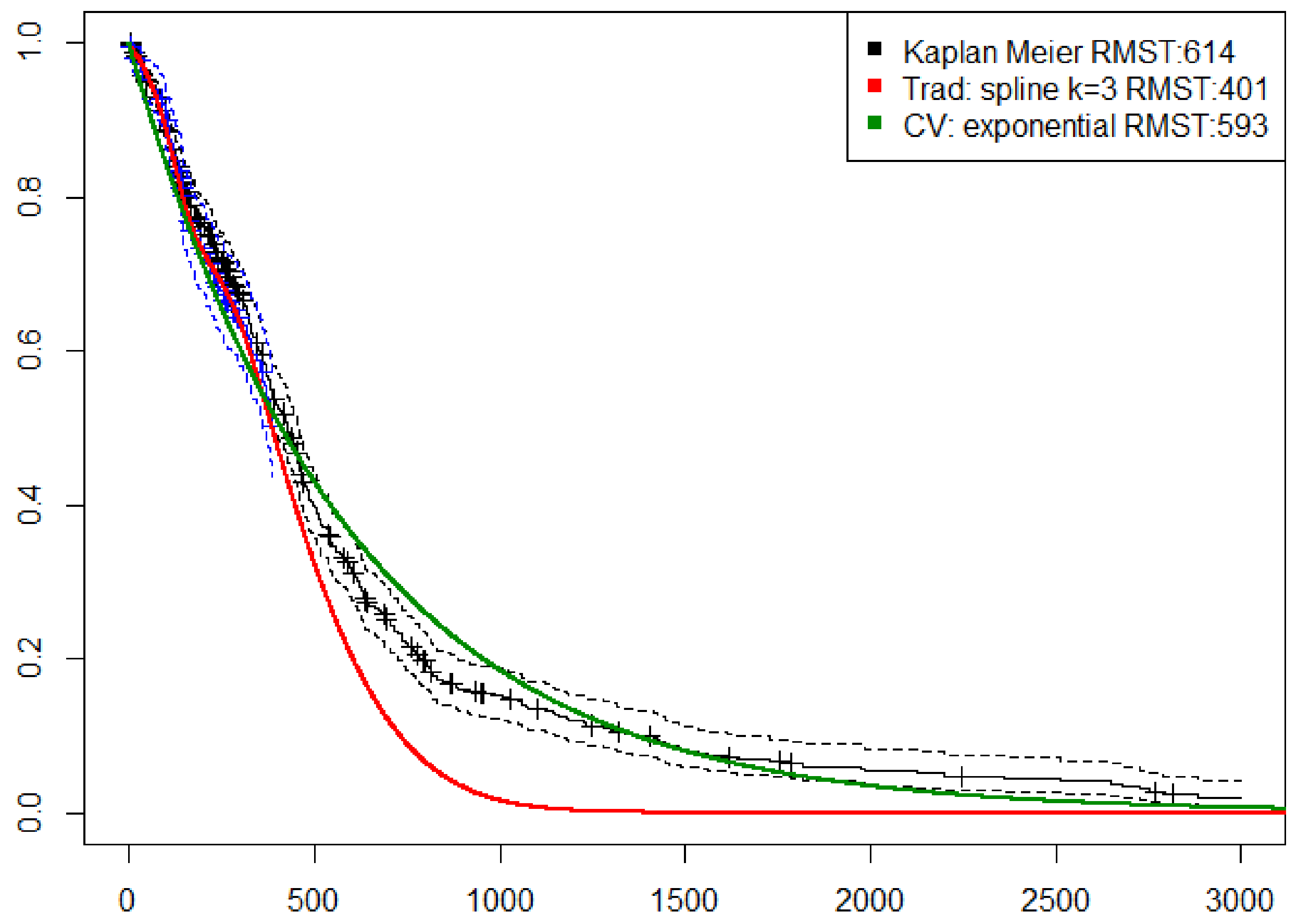


Figure 2: Example of curve fitting with simulated data lock (TCGA GBM)



- On average, RMST error was 27% higher (when based on AIC) and 40% (BIC) higher using traditional model selection compared to CV.
- CV never selected complex models (3+ parameters) while the traditional method selected complex models in 51% (AIC) and 12% (BIC) of simulations.

Table 1: RMST error (KM vs best-fitting curve to simulated data lock)

Dataset	AIC-guided selection			BIC-guided selection		
	Traditional (95% CI)	k-fold CV (95% CI)	Diff (%)	Traditional (95% CI)	k-fold CV (95% CI)	Diff (%)
GBSG	2.7 (2.3, 3.2)	2.7 (2.3, 3.2)	**0**	2.7 (2.3, 3.2)	2.5 (2.2, 2.9)	**6**
MGUS2	7.8 (6.6, 9.1)	5.5 (4.7, 6.4)	**41**	7 (5.8, 8.3)	5.5 (4.7, 6.4)	**26**
Myeloma	13.7 (11, 16.5)	7.6 (6.6, 8.8)	**80**	8.4 (7.1, 9.8)	7.6 (6.5, 8.7)	**10**
Rotterdam	3.8 (3.3, 4.4)	3.5 (2.9, 4.2)	**9**	3.7 (3.04, 4.3)	3.6 (3, 4.2)	**3**
Transplant	23.4 (21.5, 25.3)	21.1 (20, 22.2)	**11**	21.2 (20, 22.4)	21.1 (20, 22.2)	**0**
Ovarian	45.2 (43.1, 47.2)	47.9 (46.2, 49.7)	**-6**	46.9 (45.2, 48.7)	32.3 (30.7, 34)	**45**
TCGA GBM	23.4 (21.7, 25.2)	6 (5.2, 6.8)	**292**	20.1 (17.8, 22.4)	6 (5.1, 6.9)	**235**
Mean	17.2	13.5	**27**	15.7	11.2	**40**

Table 2: % of times each curve was the best fit

Model	AIC-guided selection		BIC-guided selection	
	Traditional	k-fold CV	Traditional	k-fold CV
Exponential	19	79	40	98
Gamma	3	0	5	0
Generalised gamma	5	0	3	0
Gompertz	9	1	8	0
Loglogistic	5	7	8	1
Lognormal	8	7	19	0
Splines	47	0	8	0
Weibull	5	0	2	0

Conclusions

- k-fold CV for survival model selection can regularly outperform traditional methods.
- k-fold CV favours simpler models compared to traditional methods, which may hint at their better generalisability.
- We conclude that k-fold CV may be an important addition to the modeller's toolbox when performing survival analysis.
- Further research should explore whether these findings hold in additional settings.

DISCOVER DSI

CONTACT US
☎ +32 11 26 82 98
dsi@uhasselt.be
www.UHasselt.be/DSI
Agoralaan gebouw D
3590 Diepenbeek
Belgium

DSI
DATA SCIENCE INSTITUTE
UHASSELT

Maastricht UMC+