Shrinkage of Regression Coefficients: Old Concepts and New Ideas

Daniela Dunkler *

Joint work with Georg Heinze & Willi Sauerbrei

* *Section for Clinical Biometrics*
Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna
Austria

Setting the Scene …

- **Prognostic model**
  1. Variable selection
  2. Estimation

  } use the same data

  - Large sample size ✓
  - Small to medium sample size

    ➡ overestimation of some effects

    ➡ one possible solution:

    post-estimation shrinkage

    may improve predictive accuracy

Setting the Scene …

- **Prognostic model**
  1. Variable selection
  2. Estimation

  } use the same data

  3. Estimate shrinkage factor $\hat{c}$

  4. Apply shrinkage factor to final model: $\hat{y} = \hat{c}\hat{\beta}$

Cox proportional hazards model

$$h_i(t) = h_0(t)exp(\beta_1 x_{1i} + \ldots + \beta_k x_{ki}) \qquad \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,k \end{matrix}$$

Global Shrinkage

- One common shrinkage factor for all predictors
- ≙ calibration slope of an internal validation

(Verweij & van Houwelingen 1993)

Global Shrinkage

- One common shrinkage factor for all predictors
- Leave-one-out cross-validation
  1. Re-estimate the final model n times, each time leaving out one subject (i) : $\hat{\beta}^{(-i)}$
  2. Cross-validated prognostic indices $\hat{\eta}_i^* = \sum_{j=1}^{k} \hat{\beta}_j^{(-i)} x_{ij}$
  3. $h_i(t) = h_0(t) exp(c\hat{\eta}_i^*)$

     $\hat{c}$ = global shrinkage factor

➡ globally shrunken $\hat{\gamma}_{G,j} = \hat{c}\hat{\beta}_j, \qquad j=1,\ldots,k$

(Verweij & van Houwelingen 1993)

Thrombosis Study

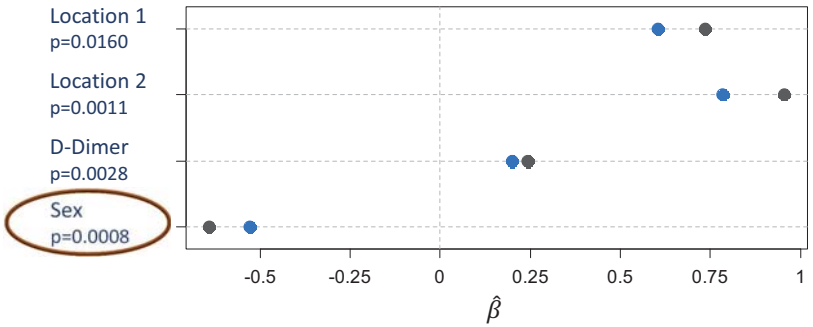
- Predict recurrence of thrombosis
- 832 patients, 143 events
- Median follow-up 43 months
- 9 candidate predictors

1. Use variable selection to derive a parsimonious model
   - Location of 1st thrombosis (3 categories)
   - D-Dimer ($\log_2$)
   - Sex
2. Estimation (with the same data)

(Eichinger *et al.* 2010)

Thrombosis Study

	Location 1	Location 2	D-Dimer	Sex
global	0.822			



Parameterwise Shrinkage

- Shrink strong predictors and weak predictors differently
- For models derived by variable selection

(Sauerbrei 1999)

PARAMETERWISE SHRINKAGE

- Shrink strong predictors and weak predictors differently
  1. Re-estimate the final model n times, each time leaving out one subject (i) : $\hat{\beta}^{(-i)}$
  2. Cross-validated partial prognostic indices
     $\hat{\eta}^*_{ij} = \hat{\beta}_j^{(-i)} x_{ij}$, j=1,…,k
  3. $h_i(t) = h_0(t) exp\left(\sum_j^k c_j \hat{\eta}^*_{ij}\right)$

     $\hat{c}_j$ = parameterwise shrinkage factors

➡ parameterwisely shrunken $\hat{\gamma}_{P,j} = \hat{c}_j \hat{\beta}_j$

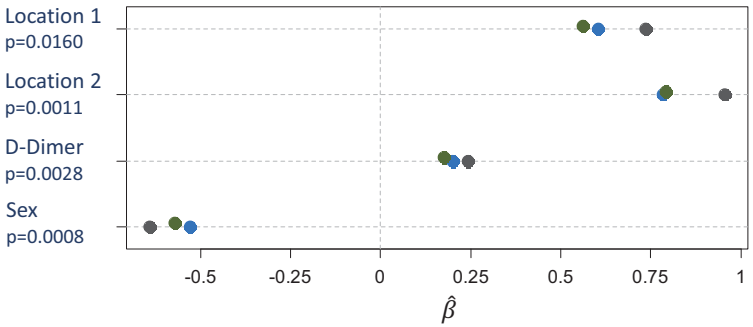
(Sauerbrei 1999)

PARAMETERWISE SHRINKAGE

- Possible to directly estimate $\hat{\gamma}_{P,j}$
  - Modify $x^*_{ij} = \hat{\eta}^*_{ij} / \hat{\beta}_j = \left(\hat{\beta}_j^{(-i)} x_{ij}\right) / \hat{\beta}_j$
  - $h_i(t) = h_0(t) exp\left(\sum_j^k \gamma_{P,j} x^*_{ij}\right)$

THROMBOSIS STUDY

	Location 1	Location 2	D-Dimer	Sex
global	0.822			
parameterwise	0.764	0.828	0.728	0.890



JOINT SHRINKAGE

- Provide one shrinkage factor for highly correlated or semantically related predictors,

  e.g.:
  - Design variables modeling a non-linear effect ($\beta_1 x + \beta_2 x^2$)
  - 2 main effects and their product term ($\beta_1 x + \beta_2 z + \beta_3 xz$)

    ➡ Coefficients cannot be interpreted separately
  - Dummy variables of a categorical predictor

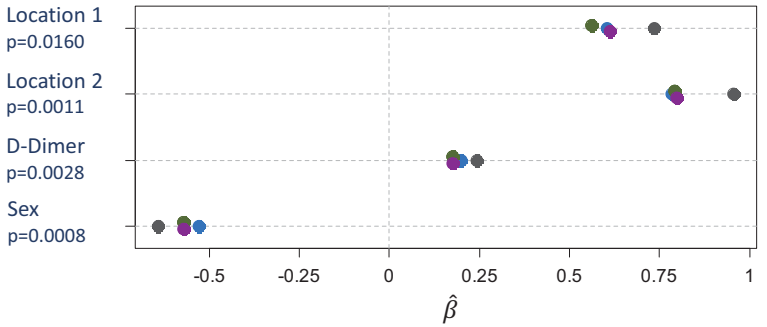
JOINT SHRINKAGE

1. Re-estimate the final model n times,
   each time leaving out one subject (i) : $\hat{\beta}^{(-i)}$
2. For β_1 to β_l: Cross-validated joint partial prognostic indices
   $\hat{\eta}^*_{i(1:l)} = \sum_{j=1}^{l} x_{ij}\hat{\beta}_j^{(-i)}$
   For $\beta_{(l+1)}$ to β_k: Cross-validated partial prognostic indices
   $\hat{\eta}^*_{ij} = x_{ij}\hat{\beta}_j^{(-i)}$
3. $h_i(t) = h_0(t)exp\left(c_1\hat{\eta}^*_{i(1:l)} + \sum_{j=l+1}^{k} c_j\hat{\eta}^*_{ij}\right)$
   → $\hat{c}_1$ = joint shrinkage factor

➡ jointly shrunken $\hat{\gamma}_{J,j} = \hat{c}_1\hat{\beta}_j$, $j=1,\ldots,l$
$\hat{\gamma}_{J,j} = \hat{c}_j\hat{\beta}_j$, $j=(l+1),\ldots,k$

THROMBOSIS STUDY

	Location 1	Location 2	D-Dimer	Sex
global	0.822			
parameterwise	0.764	0.828	0.728	0.890
joint	0.836		0.736	0.893



BREAST CANCER STUDY

- Random sample: 400 patients, 153 events
- Median follow-up 54 months
- Predict disease free survival time
- 8 candidate predictors

age	progesterone	# positive lymph nodes
tumour grade	treatment	
tumour size	menopausal status	estrogene

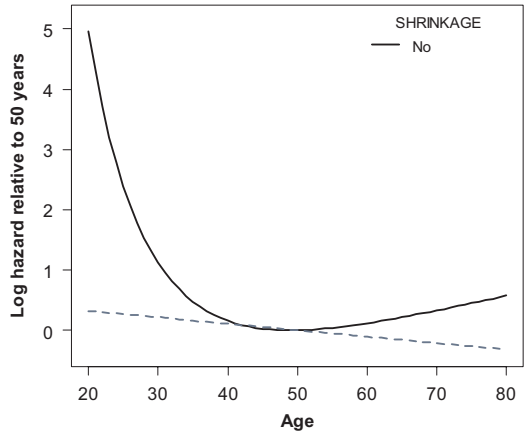
- Allow non-linear effects
  ➡ multivariable fractional polynomials

$$h_i(t) = h_0(t)exp\left(\mathbf{-5.6age^{-1} - 3.3age^{-1}log(age) -}\right.$$
$$\mathbf{-0.2prog^{0.5} - 2.3exp(-0.12nodes) +}$$
$$\left.\mathbf{+2.0grade - 0.7treat}\right)$$

(Schumacher *et al.* 1994, Sauerbrei & Royston 1999)

BREAST CANCER STUDY

$\text{Age.1} = -5.6\text{age}^{-1}$
$\text{Age.2} = -3.3\text{age}^{-1}\log(\text{age})$

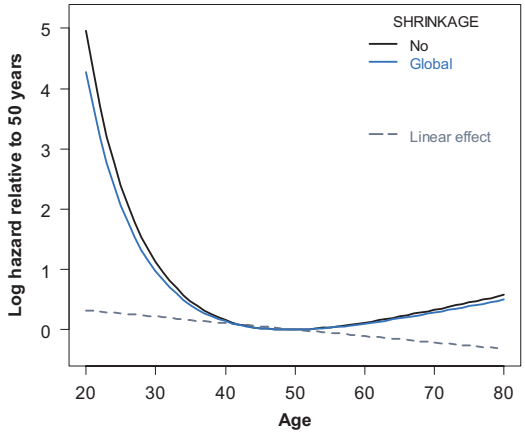


Breast Cancer Study

$\text{Age.}1 = -5.6\text{age}^{-1}$

$\text{Age.}2 = -3.3\text{age}^{-1}\log(\text{age})$

	Age.1	Age.2
G	0.862	

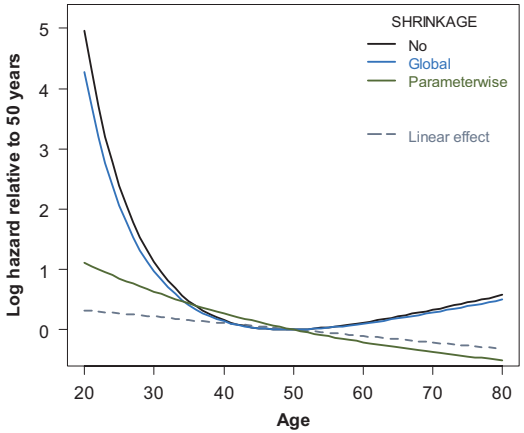


Breast Cancer Study

$\text{Age.}1 = -5.6\text{age}^{-1}$

$\text{Age.}2 = -3.3\text{age}^{-1}\log(\text{age})$

	Age.1	Age.2
G	0.862	
P	-0.232	-0.128

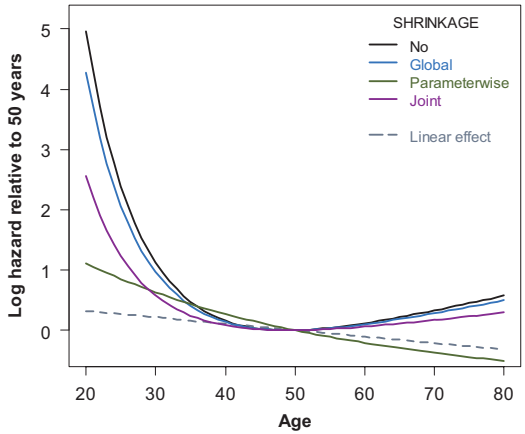


Breast Cancer Study

$\text{Age.}1 = -5.6\text{age}^{-1}$

$\text{Age.}2 = -3.3\text{age}^{-1}\log(\text{age})$

	Age.1	Age.2
G	0.862	
P	-0.232	-0.128
J	0.514	



Estimation of Shrinkage Factors

- Leave-one-out cross-validation
  - $\hat{\beta}^{(-i)}$ $i=1,\ldots,n$
  - n+2 iterative fitting processes
- Approximation with $DFBETA$ residuals
  - $\hat{\beta}^{(-i)} \approx \hat{\beta} - DFBETA_i$

+ Saves n iterative fitting processes

- Small samples: shrinkage factors may be closer to 1

Estimation of Shrinkage Factors

- Thrombosis study (n=832): joint shrinkage factors

	Leave-one-out CV	*DFBETA*	Difference
Location 1 & 2	0.8345	0.8363	**+0.22%**
D-Dimer	0.7233	0.7357	**+1.71%**
Sex	0.8896	0.8927	**+0.35%**
Computing time	6.07 sec	0.06 sec	~ 100 times

Post-Estimation Shrinkage

- Prognostic modeling: Variable selection & estimation with the same data ➞ overestimation ➞ **post-estimation shrinkage** may improve predictive accuracy

vanHouwelingen & Sauerbrei 2013

- Types of shrinkage factors:
  - Global – One common shrinkage factor

    If $0.8 < \hat{c} < 1$ ➞ $\hat{\gamma}_{G,j}$
  - Parameterwise – One shrinkage factor per predictor (prognostic relevance)

    For models derived by variable selection
  - Joint – One shrinkage factor for semantically related/highly correlated predictors (if their coefficients cannot be separated)
- Computational time: *DFBETA* approximation
- R package **`shrink`** – for linear, generalized linear, or Cox models; fractional polynomials, restricted cubic splines; availabe at cran

contact: daniela.dunkler@meduniwien.ac.at