Journal of Statistical Software

March 2016, Volume 69, Issue 8. *doi: 10.18637/jss.v069.i08*

Global, Parameterwise and Joint Shrinkage Factor Estimation

Daniela Dunkler
Medical University of Vienna

Willi Sauerbrei
University Medical Center Freiburg

Georg Heinze
Medical University of Vienna

Abstract

The predictive value of a statistical model can often be improved by applying shrinkage methods. This can be achieved, e.g., by regularized regression or empirical Bayes approaches. Various types of shrinkage factors can also be estimated after a maximum likelihood fit has been obtained: while global shrinkage modifies all regression coefficients by the same factor, parameterwise shrinkage factors differ between regression coefficients. The latter ones have been proposed especially in the context of variable selection. With variables which are either highly correlated or associated with regard to contents, such as dummy variables coding a categorical variable, or several parameters describing a nonlinear effect, parameterwise shrinkage factors may not be the best choice. For such cases, we extend the present methodology by so-called 'joint shrinkage factors', a compromise between global and parameterwise shrinkage.

Shrinkage factors are often estimated using leave-one-out resampling. We also discuss a computationally simple and much faster approximation to resampling-based shrinkage factor estimation, can be easily obtained in most standard software packages for regression analyses. This alternative may be relevant for simulation studies and other computer-intensive investigations.

Furthermore, we provide an R package **shrink** implementing the mentioned shrinkage methods for models fitted by linear, generalized linear, or Cox regression, even if these models involve fractional polynomials or restricted cubic splines to estimate the influence of a continuous variable by a nonlinear function. The approaches and usage of the package **shrink** are illustrated by means of two examples.

Keywords: global shrinkage, parameterwise shrinkage, prediction, R package, regression.

1. Introduction

Since the work of Gauss, unbiasedness of statistical estimators has been a universal paradigm for deriving statistical estimators. However, Stein (1956) and James and Stein (1961) showed

that when several parameters are simultaneously estimated, shrinking their maximum likelihood estimates towards the origin by introducing a (small) bias can reduce their mean squared error. In the context of regression modeling, it has been known since the 1970s that shrinkage methods can improve maximum likelihood estimates of regression coefficients. Hoerl and Kennard (1970a,b) introduced ridge regression, which shrinks regression coefficients towards zero by minimizing least squares penalized by the sum of squared regression coefficients. Efron and Morris (1973) explained the James-Stein estimator from an empirical Bayes perspective (Robbins 1956; Casella 1985). In 1975, Efron already asked why this improvement had not been fully embraced by statisticians and users of statistics, and even nowadays his question is still justified. From the late 1980s modern extensions of regularized likelihood estimation have been proposed including the LASSO (Tibshirani 1996, 2011), which penalizes the log likelihood by the sum of absolute regression coefficients. Greenland (2000) gave a vivid explanation why compromising unbiasedness may improve other estimation properties, and we recommend van Houwelingen (2001) for a review of shrinkage and penalization. A number of R packages (R Core Team 2015) implementing regularized and shrinkage methods are available among which the **glmnet** package (Friedman, Hastie, and Tibshirani 2010) has become particularly popular. It implements ridge, LASSO and elastic-net (a hybrid of ridge and LASSO) regularization in many types of generalized linear models. In the CRAN task view for 'Machine Learning & Statistical Learning' (Hothorn 2015) an up-to-date summary of available R packages can be found. Usually, the relative weight of the penalty term, i.e., the penalty factor, is determined by maximizing a cross-validated likelihood over a grid of penalty factors, and this connects regularized regression to empirical Bayes methods.

A simple method to shrink maximum likelihood estimates in regression models was proposed by van Houwelingen and le Cessie (1990) and Verweij and van Houwelingen (1993). They suggested to estimate a 'global' shrinkage factor by leave-one-out cross-validation (jackknifing), which is then multiplied with all regression coefficients in order to improve the fit of a model to future data. Their method was further extended by Sauerbrei (1999), who computed shrinkage factors for each regression coefficient separately, leading to 'parameterwise' shrinkage factors (PWSF). This latter approach may leave regression coefficients of explanatory variables with a strong effect nearly unshrunken, while shrinking coefficients corresponding to explanatory variables with a weaker effect. PWSFs were recommended for models obtained by variable selection (van Houwelingen and Sauerbrei 2013). Copas (1983, 1997) noted that regression coefficients of such models are on average biased away from zero and application of shrinkage methods might then be helpful. The overestimation bias arises for a variable with a weaker effect which is more likely being selected if the effect is overestimated (in absolute terms) in the particular sample (Copas and Long 1991). By means of an extensive simulation study, van Houwelingen and Sauerbrei (2013) evaluated global and parameterwise shrinkage factors in linear regression involving variable selection by backward elimination and compared them to the LASSO. They concluded that the combination of backward elimination with parameterwise shrinkage is particularly attractive, yielding prediction errors competitive with that of the LASSO, while selecting much sparser models than the LASSO. For logistic regression models, Steyerberg, Eijkemans, Harrell, and Habbema (2000) and Steyerberg, Eijkemans, and Habbema (2001) found that global shrinkage factors compared well with more sophisticated regularized methods in reducing prediction error. In general, shrinkage effects are severe in models derived by variable selection (Copas 1983), in models estimated from small data sets, or when the ratio of sample size to number of estimated coefficients is low (Steyerberg *et al.*

2000). Global and parameterwise shrinkage methods are simple alternatives to regularized methods in such situations. They are particularly attractive to practitioners because they are solely based on maximum likelihood fits and do not need specialized software. However, existence of maximum likelihood estimates is a prerequisite for their use. Therefore, if the number of explanatory variables exceeds the sample size these shrinkage methods cannot be applied, but are also not needed since shrinkage is then typically achieved by regularized methods. In the literature, shrinkage factor estimation approaches were sometimes also denoted by 'post-estimation shrinkage' and 'post-selection shrinkage' (Royston and Sauerbrei 2008; van Houwelingen and Sauerbrei 2013).

In this contribution, we extend parameterwise shrinkage for situations where some regression coefficients are naturally linked, e.g., if they correspond to dummy variables coding a categorical explanatory variable, or to two main effects and their pairwise interaction term. We propose that those regression coefficients should receive the same 'joint' shrinkage factor instead of several different parameterwise shrinkage factors. Moreover, a computationally efficient alternative to jackknifing in computing shrinkage factors using 'difference in beta' residuals (in literature and software packages commonly known as DFBETA residuals) (Belsley, Kuh, and Welsh 1980, p. 13) is proposed. Finally, we present a new R package **shrink** which can be used to estimate global, parameterwise and joint types of shrinkage factors, using the jackknife or DFBETA residuals. With the help of **shrink**, shrinkage factors can be conveniently obtained by invoking a one-line command after standard maximum likelihood estimation of linear, generalized linear, or Cox models. The package can even handle fit objects that involve fractional polynomials (FP) (Royston and Sauerbrei 2008) or restricted cubic splines (Durrleman and Simon 1989) in the predictor.

Two biomedical examples are introduced in Section 2 motivating the application of shrinkage factors. A detailed presentation of the methodology is provided in Section 3, followed by the description of the R package **shrink** in Section 4. In Section 6, multivariable regression analyses are applied to the two examples, and extended by shrinkage factor estimation using our R package **shrink**. Interpretational differences of shrinkage approaches in these examples are discussed. The paper closes with some recommendations on application of global, parameterwise and joint type shrinkage factor estimation.

2. Motivating examples

2.1. Deep vein thrombosis study

The first example deals with a prediction model for recurrence of thrombosis in individuals with deep vein thrombosis or unprovoked pulmonary embolism (Eichinger, Heinze, Jandeck, and Kyrle 2010). The data set we use here, `deepvein`, is included in the R package **shrink**, and is a slightly modified version of the original data set created to protect individual patient data. 929 individuals with a first diagnosis of venous thrombosis were followed from their date of discontinuation of oral anticoagulation for a median of 37.8 months. Time to recurrence of thrombosis was the outcome variable of primary interest, and 147 events of recurrence were observed. The following candidate predictors are available: D-dimer, which was $\log_2$-transformed (`log2ddim`), body mass index (`bmi`), duration of anticoagulation therapy (`durther`), age (`age`), sex (`sex`), location of first thrombosis (`loc`; variable with three

unordered categories; the largest group, pulmonary embolism, as reference and distal deep vein thrombosis, `loc.distal`, and proximal deep vein thrombosis, `loc.proximal`, as binary indicators), Factor II G20210A mutation (`fiimut`), and Factor V Leiden mutation (`fvleid`). Researchers preferred to have a parsimonious final model including only few predictors to facilitate clinical applicability. Using Cox regression and backward elimination the explanatory variables `log2ddim`, `sex`, and `loc` were selected for the final model.

If the same data is used for selecting explanatory variables and estimating regression coefficients, these coefficients may be overestimated and the prediction error may be underestimated. Shrinkage methods may decrease this error and strengthen external validity of the prediction model. We will revisit this example in Section 5.1, illustrating similarities and dissimilarities between global and parameterwise shrinkage factors, and we will discuss the suitability of joint shrinkage factors for nominal explanatory variables with more than two levels. We will also compare jackknife and DFBETA-type estimation of shrinkage factors.

2.2. Breast cancer study

In the second example, data from a breast cancer study (Schumacher *et al.* 1994) are used to predict disease-free survival time (death, second malignancy or cancer recurrence considered as event). We will apply shrinkage after variable selection and estimation of nonlinear influences of continuous variables using the multivariable fractional polynomial (MFP) algorithm (Royston and Sauerbrei 2008, Chapter 6), which is briefly explained in Appendix A. As explanatory variables, hormonal treatment with tamoxifen (`htreat`), and the prognostic factors age (`age`), menopausal status (`menostat`), tumor size (`tumsize`), tumor grade (`tumgrad`), number of positive lymph nodes (`posnodal`), progesterone (`prm`) and estrogen (`esm`) receptor status are available. The data comprises 686 node positive women who had complete data for these predictors; these women experienced 299 events during a median follow-up time of 53.9 months. The data is included in the R package **shrink** and is also available at `http://portal.uni-freiburg.de/imbi/Royston-Sauerbrei-book`.

Sauerbrei and Royston (1999) provide a comprehensive analysis of this data set, taking into account all prognostic factors as potential predictors. Incorporating medical knowledge into the model selection process, they proposed to represent `posnodal` by a pre-transformed variable `enodes = exp(-0.12 * posnodal)`. They considered FPs for the continuous variables `age`, `tumsize`, `prm`, `esm`, and `enodes`. In order to apply transformations with FPs, variables have to be strictly positive. Hence, `prm` and `esm` entered the model as `prm + 1` and `esm + 1`, respectively. Furthermore, `age` was divided by 50 to avoid numerical problems caused by different scalings of variables. `menostat` entered the model selection process as a binary indicator, and for `tumgrad`, which is an ordinal variable with three levels, the ordinal dummy-variable coding was used (Royston and Sauerbrei 2008, p. 56). Using a significance level of 5% Sauerbrei and Royston (1999) arrived at their preferred model III including the prognostic factors `age`, `prm`, `enodes`, and `tumgrad1`, which is the dummy variable contrasting tumor grades 2 and 3 with tumor grade 1, and `htreat`, the variable for treatment.

In Section 5.2, we will reestimate this model by applying **mfp** (Ambler and Benner 2015). However, while `htreat` was forced into model III of Sauerbrei and Royston (1999) irrespective of being significant, we now stratify the analysis by hormonal treatment. We consider this as the more appropriate approach, but because of software constraints it was not possible in the original analysis which was performed more than 15 years ago. We will demonstrate the

application of shrinkage to models with nonlinear effect estimation and selection, and discuss possible advantages of joint over parameterwise shrinkage when applied to effects of highly correlated explanatory variables.

3. Methods

3.1. Types of shrinkage factors

Assume a regression model is fitted, relating explanatory variables $X_1, \ldots, X_k$ via estimates of corresponding regression coefficients $\hat{\beta} = \left\{\hat{\beta}_1, \ldots, \hat{\beta}_k\right\}$ to some quantity of interest, such as the log odds of an event or a relative log hazard. The explanatory variables may have been selected from a larger set of candidate predictors using, e.g., backward elimination. Subject knowledge may also play an important role in this selection and estimation process, e.g., by inclusion of particular variables or interactions regardless of significance. This model development phase ends with the specification of a *final model*.

The commonly used leave-one-out (jackknife) cross-validation procedure to compute the *global* shrinkage factor c is characterized by a three-step procedure (Verweij and van Houwelingen 1993):

1. The regression coefficients of the final model are re-estimated n times, n being the number of subjects, each time leaving out one subject in turn. This will result in n vectors of regression coefficients $\hat{\beta}^{(-i)}$; $i = 1, \ldots, n$; where the superscript $(-i)$ denotes that the ith subject has been left out.

2. Cross-validated linear predictors (or 'prognostic indices') related to the n subjects can now be derived as $\eta_i = x_{i\cdot}\hat{\beta}^{(-i)}$, where $x_{i\cdot}$ denotes the covariate row vector for subject i.

3. Finally, another regression model of the same type (e.g., linear, logistic or Cox regression) and with the same dependent variable as the final model is fitted, but using $\eta = \{\eta_1, \ldots, \eta_n\}$ as the only explanatory variable. The global shrinkage factor estimate is given by the regression coefficient c estimated for η in this model, which is usually less than 1. (If the non-cross-validated prognostic indices $\eta_i = x_{i\cdot}\hat{\beta}$ were used in this model, a coefficient of 1 would be obtained.) The globally shrunken regression coefficients are given by $c\hat{\beta}_j$; $j = 1, \ldots, k$.

Sauerbrei (1999) suggested to estimate *parameterwise* shrinkage factors, modifying steps 2 and 3 of the above procedure as follows:

2. Instead of a single prognostic index η_i, now partial prognostic indices are computed as $\eta_{ij} = x_{ij}\hat{\beta}_j^{(-i)}$, $j = 1, \ldots, k$.

3. A regression model of the same type and with the same dependent variable as the final model is fitted, but using $\eta_{\cdot j} = \{\eta_{1j}, \ldots, \eta_{nj}\}$, $j = 1, \ldots, k$, as explanatory variables. This will yield PWSFs $c_1, \ldots, c_k$ related to $\eta_{\cdot 1}, \ldots, \eta_{\cdot k}$, respectively. The parameterwisely shrunken versions of the regression coefficients are $\hat{\gamma}_j = c_j\hat{\beta}_j$; $j = 1, \ldots, k$.

Type of shrinkage factor	Cross-validated prognostic indices used		
Global	η_i	$=$	$\sum_{j=1}^{k} x_{ij}\hat{\beta}_j^{(-i)}$
Parameterwise	η_{ij}	$=$	$x_{ij}\hat{\beta}_j^{(-i)}$
Joint	η_{ig}	$=$	$\sum_{j\in J_g} x_{ij}\hat{\beta}_j^{(-i)}$

Table 1: Computation of cross-validated prognostic indices η for different types shrinkage factors. Parameterwise shrinkage factors are obtained for variables $1,\ldots,k$; joint shrinkage factors for groups of variables $1,\ldots,h$; $h < k$.

Shrunken regression coefficients can also be directly obtained from a 'postfit regression model' by using as explanatory variables $x_{ij}^{*} = \eta_{ij}/\hat{\beta}_j = x_{ij}\hat{\beta}_j^{(-i)}/\hat{\beta}_j$. However, in models with an intercept (e.g., linear or generalized linear models) this requires that η_{ij}, $\hat{\beta}_j$ and $\hat{\beta}_j^{(-i)}$ were estimated with centered explanatory variables.

Explanatory variables are sometimes represented by multiple design variables, e.g., two or more dummy variables representing a categorical variable, two explanatory variables and their pairwise interaction term, or several different transformations of an explanatory variable allowing to estimate the influence of a continuous variable by a nonlinear function.These design variables are associated with regards to content and are often highly correlated. The regression coefficient of one such design variable may lack intuitive interpretation. In these cases, PWSFs of single regression coefficients are not interpretable, and their application may lead to extreme modifications of the shrunken regression coefficients compared to the original ones. This may sometimes even result in a reversed sign of a particular regression coefficient. To avoid such issues we suggest a compromise between global and parameterwise shrinkage, termed 'joint' shrinkage, which embraces these associated regression coefficients by providing one common shrinkage factor for them. More formally, assume that the indices of explanatory variables $\{1,\ldots,k\}$ are accordingly joined, yielding mutually exclusive and collectively exhaustive sets $J_1,\ldots,J_h$. Each such set J_g $(g = 1,\ldots,h)$ may correspond to one or several explanatory variables.

To compute *joint* shrinkage factors we now modify steps 2 and 3 as follows:

2. Joint partial prognostic indices are computed as $\eta_{ig} = \sum_{j\in J_g} x_{ij}\hat{\beta}_j^{(-i)}$, $g = 1,\ldots,h$.

3. A regression model of the same type and with the same dependent variable as the final model is fitted, but using $\eta_{.g} = \{\eta_{1g},\ldots,\eta_{nh}\}$, $g = 1,\ldots,h$, as explanatory variables. This will yield joint shrinkage factors $c_1,\ldots,c_h$ related to $\eta_{.1},\ldots,\eta_{.h}$, respectively. For each J_g, $g = 1,\ldots,h$, the jointly shrunken versions of the associated regression coefficient(s) $\hat{\beta}_j$ $(j \in J_g)$ are $\hat{\gamma}_j = c_g\hat{\beta}_j$.

The three types of shrinkage factors are contrasted in Table 1.

Covariance matrix of shrinkage factors

An estimate of the covariance matrix of the shrinkage factors is 'automatically' provided by the regression analysis of step 3. This covariance matrix is conditional on the estimated regression coefficients $\hat{\beta}$. Thus, it does not fully reflect the uncertainty in the estimates of the shrinkage factors, which is a very general problem of parameter estimates of models derived

data-dependently (Breiman 1992). However, it serves the researcher as a rough guide to judge the variability and correlation of the shrinkage factors. In particular, a high standard error of a shrinkage factor indicates that the shrinkage effect is not estimated precisely, and then use of the shrinkage factor might not improve the predictive value of the model for future data. A high absolute correlation between PWSFs may further indicate unstable shrinkage factor estimation, which motivates the use of a joint shrinkage factor for the involved effects. We exemplify interpretation of variability and correlation of shrinkage factors in Section 5.2.

3.2. Estimation of shrinkage factors

The above-described three-step jackknife method for estimating global, parameterwise or joint shrinkage factors requires $n + 2$ fitting processes. In logistic and Cox models, model fitting usually needs several iterations. Here we describe an approximation to the jackknife based on DFBETA residuals which in these models greatly reduces the computational burden by using quantities that are readily available after iterative model fitting.

In maximum likelihood estimation, parameter estimates $\hat{\beta}$ are commonly estimated by Newton-Raphson iteration, i.e., by iteratively updating $\hat{\beta}^{\{s+1\}} = \hat{\beta}^{\{s\}} + I^{-1}(\hat{\beta}^{\{s\}})U(\hat{\beta}^{\{s\}})$ until convergence, where the superscript $\{s\}$ refers to the sth iteration. $I(\beta)$ and $U(\beta)$ denote the observed information matrix (equalling minus the matrix of second derivatives of the log likelihood), and the score vector (the vector of first derivatives), respectively. The score vector $U(\beta)$ can be decomposed into subject-specific contributions $U_i(\beta)$, such that $\sum_i U_i(\beta) = U(\beta)$. The quantities $I(\beta)$ and $U_i(\beta)$, $i = 1, \ldots, n$, are usually available at each step of the iterative procedure.

Suppose that subject i is omitted from the analysis. Then, the parameter vector β can be re-estimated, starting from the maximum likelihood estimate for the full data set $\hat{\beta}$, by iteratively solving $\hat{\beta}^{(-i)\{s+1\}} = \hat{\beta}^{(-i)\{s\}} + I^{-1}(\hat{\beta}^{(-i)\{s\}}) \sum_{r \neq i} U_r(\hat{\beta}^{(-i)\{s\}})$ until convergence. The difference of $\hat{\beta}^{(-i)\{1\}}$ and $\hat{\beta}$ $(= \hat{\beta}^{(-i)\{0\}})$ is given by $I^{-1}(\hat{\beta}) \sum_{r \neq i} U_r(\hat{\beta})$. Since at the maximum likelihood estimate $U(\hat{\beta}_j) = 0$ for all $j = 1, \ldots, k$, $U_i(\hat{\beta}) = -\sum_{r \neq i} U_r(\hat{\beta})$. Thus, $\text{DFBETA}_i = I^{-1}(\hat{\beta})U_i(\hat{\beta})$ may serve as a one-step approximation to $\hat{\beta} - \hat{\beta}^{(-i)}$. In the sequel its k components are denoted as DFBETA_{ij}.

The jackknife-estimated $\hat{\beta}^{(-i)}$ (see Table 1) can be approximated by $\hat{\beta} - \text{DFBETA}_i$. This replacement, termed 'DFBETA method', saves n estimation procedures.

In small samples DFBETA residuals may underestimate the influence of some highly influential observations on the parameter vector, as these subjects may also affect the observed information matrix, which is ignored in the DFBETA residual approach (Therneau and Grambsch 2000). Therefore, a slight underestimation of shrinkage may be encountered when using the DFBETA method, i.e., DFBETA shrinkage factors are always slightly closer to 1 compared to their jackknife counterparts. Nevertheless, for sufficiently large data sets where individual observations have only little influence on the observed information matrix, DFBETA and jackknife values will approximately agree.

4. The R package shrink

For a given model our R package **shrink** allows to estimate global, parameterwise and joint shrinkage factors corresponding to the parameter estimates of the regression coefficients. The

package's main function `shrink` has four arguments:

- `fit` specifies the model object for which shrinkage factors are to be computed. The class of `fit` must be one of `lm`, `glm`, `coxph`, or `mfp`. The fit objects must have been called with `x = TRUE` (and `y = TRUE` when `fit` is an object of class `lm`).
- `type` specifies the type of shrinkage, and must be either `"global"`, `"parameterwise"`, or `"all"`.
- `method` specifies the shrinkage method, and must be either `"jackknife"` for shrinkage factor computation based on leave-one-out resampling, or `"dfbeta"` for its approximation by DFBETA residuals.
- If `type = "parameterwise"` or `"all"`, an additional optional argument `join` invokes the option to compute joint shrinkage factors for sets of variables. Each such set of variables can be represented by a character vector of variable names, and several sets are specified by collecting corresponding vectors in a list object. Any variable not represented in this list forms its own set.

The output of `shrink` is an object of class `shrink` with the following main components:

- `ShrinkageFactors`: A scalar value of the global shrinkage factor (if `type = "global"`), or a vector of shrinkage factors (if `type = "parameterwise"`).
- `ShrinkageFactorsVCOV`: The covariance matrix of the shrinkage factors.
- `ShrunkenRegCoef`: The shrunken regression coefficients.

An additional `predict` method provided in the **shrink** package facilitates predictions with shrunken regression coefficients. Further methods for objects of class `shrink` are `coef`, which extracts shrunken regression coefficients, `vcov`, which extracts the variance-covariance matrix of the shrinkage factors, `summary`, and `print`. The following examples illustrate application of the R package **shrink**.

5. Examples revisited

5.1. Deep vein thrombosis study

Types of shrinkage factors

Fitting the 'full' model with all available predictors by applying the `coxph` function from the **survival** package

```
R> fitfull <- coxph(Surv(time, status) ~ log2ddim + bmi + durther + age +
+    sex + loc + fiimut + fvleid, data = deepvein, x = TRUE)
```

reveals significant effects (Wald tests, significance level 0.05) of `log2ddim` ($p = 0.010$), `sex` ($p = 0.009$) and `loc` ($p = 0.014$, 2 degrees of freedom [df], expressed by `loc.distal` and

loc.proximal). The global shrinkage factor of the full model is only 0.607, and this could be taken as an indicator that the model is not useful for prediction. Estimating the nine PWSFs for this model illustrates that the PWSF approach is not adequate for such an overfitted model. Five of these PWSFs have a negative sign, indicating that even the sign in the full model may be wrong. These are strong indications that some of the variables have no or at most a weak effect and should be eliminated from the model.

Using backward elimination with an Akaike's information criterion for variable selection, which corresponds approximately to a significance level of 0.157, one arrives at a final prediction model including log2ddim ($p = 0.010$), sex ($p = 0.008$), and loc ($p = 0.010$, 2 df). The unshrunken regression coefficients of this model can be estimated by invoking step:

```
R> fitd <- step(fitfull, direction = "backward")
```

From model fitd, shown below, we learn that high D-dimer, male sex and a first pulmonary embolism (which is the reference group for loc.distal and loc.proximal) are associated with earlier recurrence. The final model can be displayed by print(fitd):

```
               coef exp(coef) se(coef)     z      p
log2ddim      0.219     1.245   0.0854  2.56 0.0100
sex.male      0.491     1.634   0.1847  2.66 0.0079
loc.distal   -0.922     0.398   0.3101 -2.97 0.0029
loc.proximal -0.205     0.815   0.1787 -1.15 0.2500

Likelihood ratio test=24.5 on 4 df, p=6.37e-05, n=929, number of events=147
```

The global shrinkage factor, quantifying the required correction for overestimation, and shrunken regression coefficients can be obtained using the **shrink** package with a simple one-line command:

```
R> shrink(fitd, type = "global", method = "jackknife")

Shrinkage Factors (type=global, method=jackknife):
[1] 0.808
Shrunken Regression Coefficients:
log2ddim  sex.male  loc.distal  loc.proximal
0.177     0.396     -0.745      -0.166
```

The global shrinkage factor of the final model is 0.808, and is remarkably higher than the global shrinkage factor of the unselected model (0.607), but still indicates some overestimation. This residual overestimation may be removed by multiplying the regression coefficients with the global shrinkage factor, leading to predictions that are shrunken towards the mean expected outcome over all patients.

A large patient-level meta-analysis revealed that sex is the strongest predictor of recurrence (Douketis *et al.* 2011). This medical knowledge indicates that the inclusion of sex in our model is probably not an accidental result caused by chance and may not need much correction for selection bias. Consequently, its globally shrunken regression coefficient may underestimate its true relevance for prediction. PWSFs were suggested to overcome this problem, and to assign different shrinkage factors to each explanatory variable. In our example, PWSFs can be obtained by the command:

```
R> pwsf <- shrink(fitd, type = "parameterwise", method = "jackknife")

Shrinkage Factors (type=parameterwise, method=jackknife):
log2ddim  sex.male  loc.distal  loc.proximal
0.732     0.835     0.839       0.132

Shrunken Regression Coefficients:
log2ddim  sex.male  loc.distal  loc.proximal
0.160     0.410     -0.774      -0.027
```

The four PWSFs reflect the unequal importance of shrinkage for the four design variables for prediction, ranging from 0.132 for `loc.proximal` to 0.839 for `loc.distal`.

The approximate covariance matrix for these shrinkage factors can be assessed by `pwsf$ShrinkageFactorsVCOV`. In our case, the standard error for the PWSF of `locproximal` is estimated at 0.870, being considerably larger then the standard errors of the PWSFs of the other design variables in the model.

```
R> sqrt(diag(pwsf$ShrinkageFactorsVCOV))

log2ddim  sex.male  loc.distal  loc.proximal
0.370     0.370     0.327       0.870
```

When associated dummy variables such as `loc.distal` and `loc.proximal` are jointly selected, joint shrinkage of the corresponding coefficients may be more appropriate. In our example, we obtain a joint shrinkage factor of 0.806 with the command:

```
R> shrink(fitd, join = list(c("loc.distal", "loc.proximal")))

Shrinkage Factors (type=parameterwise with join option, method=jackknife):
log2ddim  sex.male  loc.distal  loc.proximal
0.781     0.836     0.806       0.806

Shrunken Regression Coefficients:
log2ddim  sex.male  loc.distal  loc.proximal
0.171     0.411     -0.743      -0.165
```

Jackknife and DFBETA estimation of shrinkage factors

Now we turn to the estimation method of shrinkage factors, which could be by jackknife (Section 3.1) or by DFBETA approximation (Section 3.2). In the example above we have only used the jackknife (`method = "jackknife"`); the DFBETA method is invoked by specifying `method = "dfbeta"`. Table 2 compares the results from both methods. Overall, the DFBETA shrinkage factors exceed their jackknife counterparts by maximally 0.015. However, while the call to `shrink` with the jackknife method took 3.03 seconds on a contemporary PC, the DFBETA method only required 0.02 seconds which is about 150 times faster. As long as a single model is evaluated, this does not make much difference in absolute time, but time savings multiply when using computer-intensive methods, such as, evaluating model stability or performing model validation by resampling methods or in simulation studies (Sauerbrei 1999).

Explanatory variable	Jackknife	DFBETA	Relative difference
Global shrinkage	0.8076	0.8123	0.6%
Parameterwise shrinkage			
`log2ddim`	0.7321	0.7385	0.9%
`sex.male`	0.8351	0.8373	0.3%
`loc.distal`	0.8394	0.8449	0.7%
`loc.proximal`	0.1321	0.1470	11.2%
Joint shrinkage			
`log2ddim`	0.7806	0.7864	0.7%
`sex.male`	0.8364	0.8386	0.3%
`loc`	0.8055	0.8111	0.7%
Computing time	3.03	0.02	-99.3%

Table 2: Deep vein thrombosis study: Comparison of jackknife and DFBETA-approximated shrinkage factors. Computing time is given for parameterwise shrinkage factors in seconds.

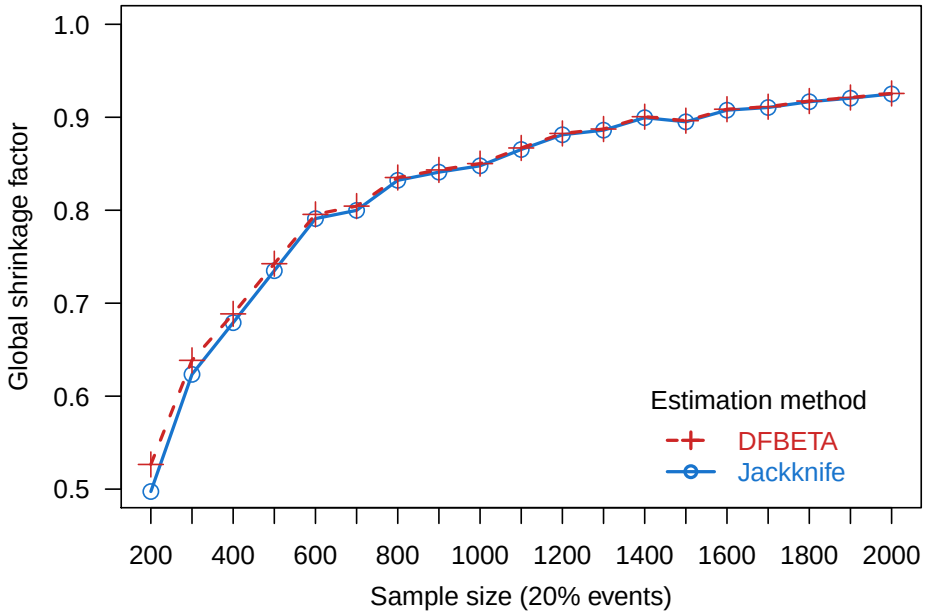

Figure 1: Data resampled from deep vein thrombosis study: Comparison of jackknife and DFBETA-approximated global shrinkage factors.

By means of simulating data from the deep vein thrombosis study with various sample sizes, we demonstrate the typical deviation of DFBETA and jackknife methods depending on sample size and magnitude of shrinkage factors (Figure 1). Specifically, we resampled 200–1400 subjects from the original data set with replacement, fitted models with `log2ddim`, `sex`, and `loc` as explanatory variables, and computed the global shrinkage factors. This was repeated 100 times, and the median values of the shrinkage factors at each sample size are

shown in Figure 1. In summary, this investigation led us to the following conclusions:

- *Sample size*: The deviation of DFBETA from jackknife shrinkage factors depends on the sample size (more deviation with smaller samples). Note that the estimated regression coefficients have equal average values across sample sizes, but considerably higher variance in smaller samples compared to larger ones. This causes the lower shrinkage factors with smaller samples.
- *Magnitude of shrinkage factors*: The smaller the shrinkage factors, the higher is the deviation of DFBETA from jackknife shrinkage factors. If shrinkage factors are not too low (> 0.5), then the deviations of jackknife and DFBETA can be expected to be less than 10%.
- *Time savings* by the DFBETA method compared to the jackknife method are substantial and grow exponentially with increasing sample size.

5.2. Breast cancer study

In this example, we will apply shrinkage after variable selection and estimation of nonlinear effects using the MFP algorithm. Following the approach of Sauerbrei and Royston (1999) with the only exception that the model is stratified by `htreat`, we reestimate their model III by using the **mfp** package (Ambler and Benner 2015). Model selection (at a significance level of 5%) and estimation can be performed by submitting the following code:

```
R> contrasts(GBSG$tumgrad) <- matrix(c(0, 1, 1, 0, 0, 1), ncol = 2,
+    dimnames = list(1:3, c("tumgrad1", "tumgrad2")), byrow = FALSE)

  tumgrad1  tumgrad2
1        0         0
2        1         0
3        1         1

R> fitg <- mfp(Surv(rfst, cens) ~ fp(age) + fp(prm) + fp(esm) +
+    fp(tumsize) + fp(enodes) + tumgrad + menostat + strata(htreat),
+    family = cox, data = GBSG, alpha = 0.05, select = 0.05)
```

The backward elimination applied by `mfp` with a significance level of 0.05 selects `age`, `prm`, `enodes`, and `tumgrad1`, the dummy variable contrasting tumor grade 1 from grades 2 and 3. `mfp` applies a built-in pretransformation algorithm which gently transforms variables by shifting them to the positive range, if necessary, and by approximately aligning their scalings (Royston and Sauerbrei 2008, Chapter 4.11). Here, `mfp` determines pretransformations (`PT(·)`) for `age` and `prm` of `PT(age) = age / 100` and `PT(prm) = (prm + 1) / 100`, respectively. For `PT(age)`, `mfp` selects an FP of degree 2 (powers of -2 and -1), leading to the design variables `age.1 = PT(age) ^-2` and `age.2 = PT(age) ^-1`. Furthermore, `mfp` suggests a square root transformation for `PT(prm)` (`prm.1 = PT(prm) ^0.5`), and a linear function for `enodes`. The estimated regression coefficients of the selected model are displayed by `print(fitg)`:

Explanatory variable	Design variable	$\hat{\beta}_j$	df	p value	Parameterwise shrinkage (SE)	Joint shrinkage (SE)
age			4	0.001		0.876 (0.188)
	age.1	0.606			0.811 (0.236)	
	age.2	−2.654			0.782 (0.277)	
prm	prm.1	−0.572	4	<0.001	0.978 (0.189)	0.982 (0.189)
posnodal	enodes.1	−1.978	4	<0.001	0.987 (0.116)	0.986 (0.116)
tumgrad1	tumgrad1	0.514	1	0.028	0.811 (0.453)	0.809 (0.452)

Table 3: Breast cancer study: Regression coefficients $\hat{\beta}_j$, degrees of freedom (df), p values, parameterwise, and joint shrinkage factors with respective standard errors (SE) based on the jackknife method.

```
              coef exp(coef) se(coef)      z        p
age.1       0.6060   1.83303   0.1188  5.101 3.37e-07
age.2      -2.6539   0.07037   0.5902 -4.496 6.91e-06
prm.1      -0.5717   0.56456   0.1109 -5.154 2.55e-07
enodes.1   -1.9782   0.13832   0.2272 -8.706 0.00e+00
tumgrad1    0.5137   1.67141   0.2495  2.059 3.95e-02

Likelihood ratio test=142 on 5 df, p=0 n=686
```

The p values given in the output above, based on 1 df, refer to the design variables, and are thus conditional on the selection. In Table 3, we supply p values based on likelihood ratio tests for the explanatory variables which are unconditional of the selected degree and power.

The following two lines estimate global and parameterwise shrinkage factors:

```
R> globalsf <- shrink(fitg, type = "global", method = "jackknife")
R> pwsf <- shrink(fitg, type = "parameterwise", method = "jackknife")
```

The global shrinkage factor is 0.953 with a standard error of 0.081. PWSFs of `prm.1` and `enodes.1` are above 0.975, and `tumgrad1`'s PWSF is somewhat lower with 0.811. The PWSFs for `age.1` and `age.2` are 0.811 and 0.782, respectively. The corresponding SEs are large in absolute terms, ranging from 0.116 for `enodes.1` to 0.453 for `tumgrad1`. As outlined in Section 3.1, regression coefficients and PWSFs of `age.1` and `age.2` are not interpretable, as the two effects are components of one function estimating the influence of the explanatory variable `age` on the outcome. This is also reflected in the PWSFs' estimated correlation matrix, which can be obtained by submitting the command `cov2cor(pwsf$ShrinkageFactorsVCOV)`:

```
            age.1   age.2   prm.1   enodes.1  tumgrad1
age.1       1.000   0.984   0.026   0.030    -0.040
age.2       0.984   1.000   0.032   0.021    -0.035
prm.1       0.026   0.032   1.000  -0.055    -0.200
enodes.1    0.030   0.021  -0.055   1.000    -0.078
tumgrad1   -0.040  -0.035  -0.200  -0.078     1.000
```

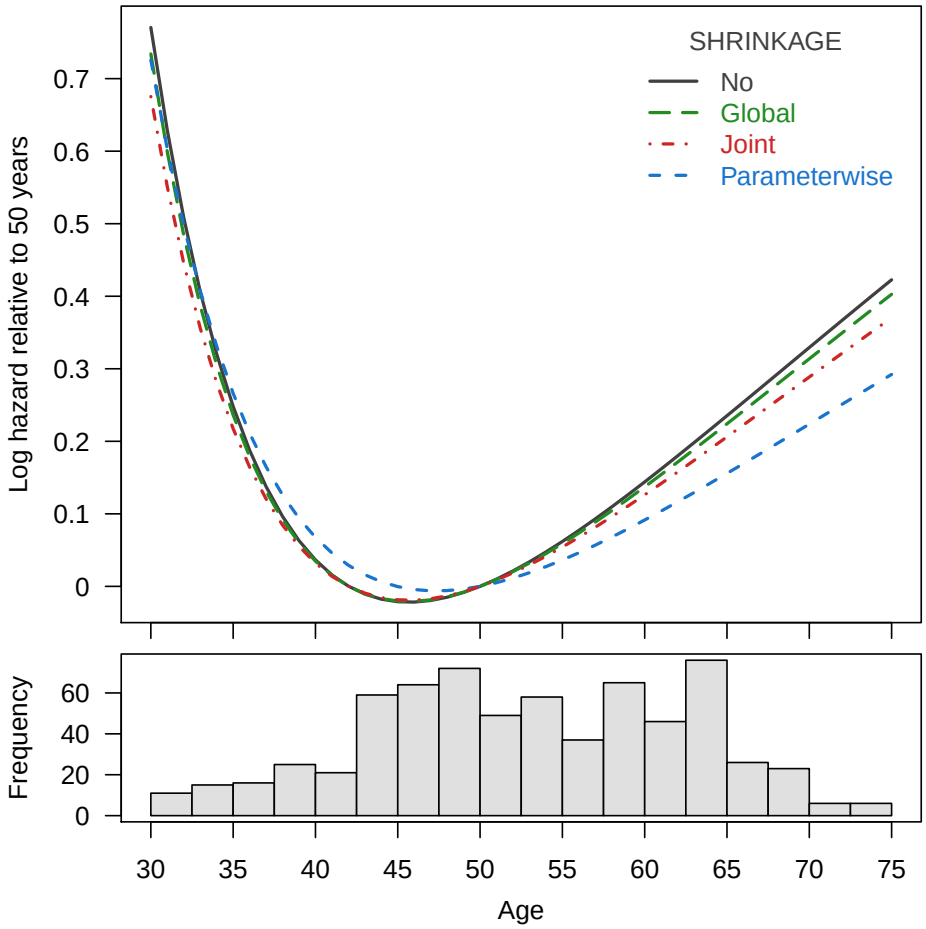

Figure 2: Breast cancer study: Estimates of the log hazard of age relative to 50 years before and after post-estimation shrinkage (top) and frequency distribution of `age` (bottom).

All correlations between PWSFs of design variables from different explanatory variables are low. Only the correlation of the PWSFs of `age.1` and `age.2` is close to 1, which further motivates the use of a joint shrinkage factor for these effects:

```
R> shrink(fitg, type = "parameterwise", method = "jackknife",
+    join = list(c("age.1", "age.2")))
```

The joint shrinkage factor of `age` of 0.876 is larger than the PWSFs of the individual components `age.1` and `age.2`, while the standard error (0.188) is lower.

In Figure 2 we have compared the resulting functional form of the `age` effect on the log hazard scale when applying no, global, parameterwise or joint shrinkage factor estimation. The selected functional form for `age` suggests a U-shaped risk profile, with the lowest risk for women aged around 46 years. This functional form determined by estimation with FPs is not affected by global or joint shrinkage, our preferred method in this example. However, parameterwise shrinkage changes the functional form and the magnitude of the effect of `age` particularly for older women.

6. Concluding remarks

We discussed three types of shrinkage factors obtained after maximum likelihood estimation, which quantify by how much regression coefficients may need to be corrected for overestimation. A global shrinkage factor penalizes all predictors equally. By contrast, PWSFs shrink weak, less relevant predictors more than strong ones, and in the extreme case, a PWSF may be close to 0 indicating 'random selection' of a variable and its irrelevance for prediction. The PWSFs of regression coefficients related to a set of associated design variables may be misleading, as the effects of these design variables are inseparable. Hence, the corresponding regression coefficients should be shrunken jointly using the proposed joint shrinkage factor. In general, joint shrinkage should replace parameterwise shrinkage in situations where a single regression coefficient β_j is uninterpretable because of the inherent dependency of the design variable corresponding to β_j on other design variables.

In Section 5.2 we used MFP as it is a well-described algorithm that simultaneously selects variables and estimates their functional form. A researcher could also estimate nonlinear effects of continuous explanatory variables using restricted cubic splines, which are also supported by **shrink**. Similarly as in the case of fractional polynomials, the individual components of restricted cubic splines have no interpretation and therefore joint shrinkage factors should be applied.

One should also consider joint shrinkage when multiple dummy variables of a categorical explanatory variable were selected jointly in a significance-based selection procedure. Parameterwise shrinkage factors may still be applied if dummy variables are separately tested and selected. This latter approach may sometimes lead to a more parsimonious model, since it allows for a data-driven collapsing of some categories with the reference category, but it requires careful selection of a suitable reference. For a deeper discussion on coding and selection of categorical variables, we refer to Royston and Sauerbrei (2008, Chapter 3).

Usually global, parameterwise or joint shrinkage factors are estimated for the purpose of decreasing the prediction error of a model, but it may not be clear upfront which type of shrinkage factor is best suited for this purpose. In addition to the magnitude of the shrinkage factors, their covariance matrix can help the researcher with this decision. PWSFs which are much lower than 1, and have high correlations or large variances may severely inflate the prediction error (van Houwelingen and Sauerbrei 2013). This is likely the case when variables are correlated.

To avoid this unfavorable consequence of PWSFs, a researcher could estimate a joint shrinkage factor for each set of correlated variables or variables related by content. For example, in an epidemiological study variable groups could result from separating environmental exposures ('above the skin' exposures) from variables describing the genetic or clinical state of individuals ('below the skin' variables). Sometimes, the joint shrinkage factors are still correlated and low, and then they may be better replaced by a single global shrinkage factor, estimated with lower variance.

We have also presented a new method for shrinkage factor estimation, based on DFBETA residuals, as an alternative to the more time-consuming leave-one-out (jackknife) method. This variant may be of particular interest in resampling-based analyses or in simulation studies. In our examples, as expected, jackknife shrinkage factors were closely approximated by their DFBETA counterparts.

These shrinkage types (global, parameterwise and joint) and shrinkage methods (jackknife and

DFBETA) have been implemented in the R package **shrink** enabling convenient and efficient computation of shrinkage factors, their correlation structure and their variances with an easy-to-use one-line command. By providing methods applicable to the classes `lm`, `glm`, `coxph` and `mfp` many statistical models used in biomedical research and in other disciplines are supported. The package is available at https://CRAN.R-project.org/package=shrink or at http://cemsiis.meduniwien.ac.at/en/kb/science-research/software/. The package manual contains additional examples for all supported object classes.

Acknowledgments

This work was partly funded by the European Union (Grant Agreement No. HEALTH F2-2009-241544). We are grateful to Drs Eichinger and Kyrle, Medical University of Vienna, for providing the data of the deep vein thrombosis study. We would like to thank the handling editor and two reviewers for many valuable comments which considerably improved the manuscript and the R package **shrink**.

References

Ambler G, Benner A (2015). ***mfp**: Multivariable Fractional Polynomials*. R package version 1.5.2, URL https://CRAN.R-project.org/package=mfp.

Belsley DA, Kuh E, Welsh R (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Chichester. doi:10.1002/0471725153.

Breiman L (1992). "The Little Bootstrap and Other Methods for Dimensionality Selection in Regression: X-Fixed Prediction Error." *Journal of the American Statistical Association*, **87**(419), 738–754. doi:10.1080/01621459.1992.10475276.

Casella G (1985). "An Introduction to Empirical Bayes Data Analysis." *The American Statistician*, **39**(2), 83–87. doi:10.2307/2682801.

Copas JB (1983). "Regression, Prediction and Shrinkage." *Journal of the Royal Statistical Society B*, **45**(3), 311–354.

Copas JB (1997). "Using Regression Models for Prediction: Shrinkage and Regression to the Mean." *Statistical Methods in Medical Research*, **6**(2), 167–183.

Copas JB, Long T (1991). "Estimating the Residual Variance in Orthogonal Regression with Variable Selection." *Journal of the Royal Statistical Society D*, **40**(1), 51–59. doi:10.2307/2348223.

Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, Kyrle P, Poli D, Tait RC, Iorio A (2011). "Risk of Recurrence After Venous Thromboembolism in Men and Women: Patient Level Meta-Analysis." *British Medical Journal*, **342**, d813. doi:10.1136/bmj.d813.

Durrleman S, Simon R (1989). "Flexible Regression Models with Cubic Splines." *Statistics in Medicine*, **8**(5), 551–561. doi:10.1002/sim.4780080504.

Efron B (1975). "Biased Versus Unbiased Estimation." *Advances in Mathematics*, **16**(3), 259–277. doi:10.1016/0001-8708(75)90114-0.

Efron B, Morris C (1973). "Stein's Estimation Rule and Its Competitors – An Empirical Bayes Approach." *Journal of the American Statistical Association*, **68**(341), 117–130. doi:10.1080/01621459.1973.10481350.

Eichinger S, Heinze G, Jandeck LM, Kyrle P (2010). "Risk Assessment of Recurrence in Patients With Unprovoked Deep Vein Thrombosis or Pulmonary Embolism – The Vienna Prediction Model." *Circulation*, **121**(14), 1630–1636. doi:10.1161/circulationaha.109.925214.

Friedman JH, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, **33**(1), 1–22. doi:10.18637/jss.v033.i01.

Greenland S (2000). "Principles of Multilevel Modelling." *International Journal of Epidemiology*, **29**(1), 158–167. doi:10.1093/ije/29.1.158.

Hoerl AE, Kennard RW (1970a). "Ridge Regression: Applications to Nonorthogonal Problems." *Technometrics*, **12**(1), 69–82. doi:10.1080/00401706.1970.10488635.

Hoerl AE, Kennard RW (1970b). "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics*, **12**(1), 55–67. doi:10.2307/1267351.

Hothorn T (2015). "CRAN Task View: Machine Learning & Statistical Learning." Version 2015-12-02, URL https://CRAN.R-project.org/view=MachineLearning.

James W, Stein C (1961). "Estimation with Quadratic Loss." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 361–379. University of California Press, Berkeley, California.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robbins H (1956). "An Empirical Bayes Approach to Statistics." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 157–163. University of California Press, Berkeley, California.

Royston P, Altman DG (1994). "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling." *Journal of the Royal Statistical Society C*, **43**(3), 429–467. doi:10.2307/2986270.

Royston P, Sauerbrei W (2008). *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. John Wiley & Sons, Chichester.

Sauerbrei W (1999). "The Use of Resampling Methods to Simplify Regression Models in Medical Statistics." *Journal of the Royal Statistical Society C*, **48**(3), 313–329. doi:10.1111/1467-9876.00155.

Sauerbrei W, Meier-Hirmer C, Benner A, Royston P (2006). "Multivariable Regression Model Building by Using Fractional Polynomials: Description of SAS, Stata and R Programs." *Computational Statistics & Data Analysis*, **50**(12), 3464 – 3485. doi:10.1016/j.csda.2005.07.015.

Sauerbrei W, Royston P (1999). "Building Multivariable Prognostic and Diagnostic Models: Transformation of the Predictors by Using Fractional Polynomials." *Journal of the Royal Statistical Society A*, **162**(1), 72–94. doi:10.1111/1467-985x.00122.

Schumacher M, Basert G, Bojar H, Huebner K, Olschewski M, Sauerbrei W, Schmoor C, Beyerle C, Neumann RLA, Rauschecker HF (1994). "Randomized 2 × 2 Trial Evaluating Hormonal Treatment and the Duration of Chemotherapy in Node-Positive Breast Cancer Patients." *Journal of Clinical Oncology*, **12**(10), 2086–2093.

StataCorp (2011). *Stata Data Analysis Statistical Software: Release 12*. StataCorp LP, College Station. URL http://www.stata.com/.

Stein C (1956). "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 197–206. University of California Press, Berkeley, California.

Steyerberg EW, Eijkemans MJC, Habbema JDF (2001). "Application of Shrinkage Techniques in Logistic Regression Analysis: A Case Study." *Statistica Neerlandica*, **55**(1), 76–88. doi:10.1111/1467-9574.00157.

Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JD (2000). "Prognostic Modelling with Logistic Regression Analysis: A Comparison of Selection and Estimation Methods in Small Data Sets." *Statistics in Medicine*, **19**(8), 1059–1079. doi:10.1002/(sici)1097-0258(20000430)19:8<1059::aid-sim412>3.3.co;2-s.

Therneau TM, Grambsch PM (2000). *Modeling Survival Data. Extending the Cox Model*. Springer-Verlag, New York. doi:10.1007/978-1-4757-3294-8.

Tibshirani R (1996). "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society B*, **58**(1), 267–288.

Tibshirani R (2011). "Regression Shrinkage and Selection via the LASSO: A Retrospective." *Journal of the Royal Statistical Society B*, **73**(3), 273–282. doi:10.1111/j.1467-9868.2011.00771.x.

van Houwelingen HC, Sauerbrei W (2013). "Cross-Validation, Shrinkage and Variable Selection in Linear Regression Revisited." *Open Journal of Statistics*, **3**(2), 79–102. doi:10.4236/ojs.2013.32011.

van Houwelingen JC (2001). "Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy." *Statistica Neerlandica*, **55**(1), 17–34. doi:10.1111/1467-9574.00154.

van Houwelingen JC, le Cessie S (1990). "Predictive Value of Statistical Models." *Statistics in Medicine*, **9**(11), 1303–1325. doi:10.1002/sim.4780091109.

Verweij PJM, van Houwelingen HC (1993). "Cross-Validation in Survival Analysis." *Statistics in Medicine*, **12**(24), 2305–2314. doi:10.1002/sim.4780122407.

A. Fractional polynomials and MFP

Royston and Altman (1994) suggested that to account for a possibly nonlinear effect of a continuous explanatory variable x, one or two transformations of type x^p, termed fractional polynomials (FP) of degree 1 or 2, respectively, often suffice as design variables in the model. They proposed to use the following set of eight power terms $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. If $p = 0$, the logarithm of x is used. Another special case is the use of 'repeated powers', i.e., two transformations with the same power p. In this case, the second transformation is multiplied by $\log(x)$, such that the two fractional polynomial terms are defined by x^p and $x^p \log(x)$. Linear pretransformations of explanatory variables could be necessary to shift all values to the strictly positive range, which is a prerequisite for applying power exponents of less than one. There are eight FP1 functions with one power term and 36 FP2 functions with two power terms. To decide which of these FP functions is the most suitable one for a given problem, Royston and Sauerbrei (2008, Chapter 4.10.2) described a function selection procedure (FSP) which preserves the type I error probability at any desired level.

For multivariable model building requiring the selection of important variables and determination of the functional form for continuous variables the MFP algorithm has been developed (Royston and Sauerbrei 2008, Chapter 6). It combines backward elimination with the FSP. A final model produced by the MFP algorithm may either select an explanatory variable using a fractional polynomial transformation, using no transformation (effect is linear), or exclude a variable from the model. These decisions are based on significance tests at the level α. The main implementation for MFP is in Stata (StataCorp. 2011), but it is also implemented as a SAS macro (Sauerbrei, Meier-Hirmer, Benner, and Royston 2006) and in the R package **mfp** (Ambler and Benner 2015).

Affiliation:

Georg Heinze
Section for Clinical Biometrics
Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna
1090 Vienna, Austria
E-mail: georg.heinze@meduniwien.ac.at
URL: http://www.meduniwien.ac.at/user/georg.heinze/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
March 2016, Volume 69, Issue 8 *Submitted:* 2013-12-23
doi:10.18637/jss.v069.i08 *Accepted:* 2015-02-18