Laurin, C. A., Boomsma, D., & Lubke, G. (2016). The use of vector bootstrapping to improve variable selection precision in Lasso models. Statistical Applications in Genetics and Molecular Biology, 15(4), 305-320. DOI: 10.1515/sagmb-2015-0043

Charles Laurin*, Dorret Boomsma and Gitta Lubke

The use of vector bootstrapping to improve variable selection precision in Lasso models

DOI 10.1515/sagmb-2015-0043

Abstract: The Lasso is a shrinkage regression method that is widely used for variable selection in statistical genetics. Commonly, *K*-fold cross-validation is used to fit a Lasso model. This is sometimes followed by using bootstrap confidence intervals to improve precision in the resulting variable selections. Nesting cross-validation within bootstrapping could provide further improvements in precision, but this has not been investigated systematically. We performed simulation studies of Lasso variable selection precision (VSP) with and without nesting cross-validation within bootstrapping. Data were simulated to represent genomic data under a polygenic model as well as under a model with effect sizes representative of typical GWAS results. We compared these approaches to each other as well as to software defaults for the Lasso. Nested cross-validation had the most precise variable selection at small effect sizes. At larger effect sizes, there was no advantage to nesting. We illustrated the nested approach with empirical data comprising SNPs and SNP-SNP interactions from the most significant SNPs in a GWAS of borderline personality symptoms. In the empirical example, we found that the default Lasso selected low-reliability SNPs and interactions which were excluded by bootstrapping.

Keywords: additive-by-additive epistasis; association; bootstrap; Lasso; polygenic model; variable selection.

1 Introduction

Multiple linear regression is most useful when it is applied to samples in which the number of predictors is small relative to the number of observations. However, if the number of predictors is large relative to the number of observations, there will be considerable sampling variability in the estimated coefficients. Under those conditions, the increased variability of coefficients can be managed by applying shrinkage methods. Shrinkage means estimating coefficients under a constraint that leads them to have reduced absolute values, drawing them toward 0, thus reducing sampling variability. There are many such constraints, hence many shrinkage methods. One of the most important shrinkage methods is the Lasso (Tibshirani, 2011).

When the Lasso is applied to a multiple regression problem, small-valued coefficient estimates are reduced to 0 and the remaining coefficient estimates are shrunk by a fixed amount (Tibshirani, 2013). Because of this property, the Lasso is often used for variable selection (Tibshirani, 1996). In Lasso variable selection, the predictors having nonzero coefficient estimates after shrinkage are selected into the regression model, and those that are shrunk to 0 are excluded from the model.

Variable selection with the Lasso is the task of deciding whether the predictor variables that could be included in a regression model are “unimportant” or as “important.” The unimportant, or “noise,” predictors are not associated with the outcome in the population, but could be strongly associated in a given sample because of sampling fluctuation. Important predictors are consistently associated with the outcome over

***Corresponding author: Charles Laurin,** Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, BS8 2BN, UK, e-mail: cl14022@bristol.ac.uk. http://orcid.org/0000-0003-2439-9004; and Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA
Dorret Boomsma: Department of Biological Psychology, VU University Amsterdam, Amsterdam, 1081 HV, Netherlands
Gitta Lubke: Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA; and Department of Biological Psychology, VU University Amsterdam, Amsterdam, 1081 HV, Netherlands

independent samples. Unimportant predictors should be excluded from the regression model (i.e. their coefficient estimates shrunken to 0), while important predictors should be included in the model.

In practice, using the Lasso for variable selection translates to classifying the excluded predictors as unimportant and the included predictors as important. The accuracy of this claim has been studied in simulation studies, and the Lasso has often been found to have low variable selection precision (VSP), meaning that among the included predictors, a relatively large proportion (>50%) were false positives (Devlin et al., 2003; Ayers and Cordell, 2010; He and Lin, 2011). In this paper, we propose a method for controlling the Lasso's VSP.

We build on a variety of research that has investigated how to control the Lasso's VSP. A main current in this research has been to estimate standard errors (SEs) and confidence intervals (CIs) for Lasso coefficient estimates with the bootstrap (e.g. Chatterjee 2011; Minnier et al., 2011). Lasso SEs and CIs are used in a secondary variable selection step based on analogy with hypothesis testing (Freedman and Lane, 1983). The SEs are used to generate t-statistics; predictors having t-statistics below a user-set cutoff (e.g. $|t|\leq 2$) are excluded. Similarly, predictors with CIs that contain 0 are excluded.

Successfully applying a second variable selection step requires knowledge of why the Lasso includes too many variables in the model during the first step. Lasso variable selection depends on its degree of shrinkage. The degree of shrinkage in a Lasso model is controlled by a user-set metaparameter called λ.

The shrinkage metaparameter λ is a scalar; its value determines the number of variables that the Lasso selects. The larger the value of λ, the more conservative the model. The choice of λ is thus related to the Lasso's VSP. In general, each value of λ is associated with a single Lasso model. λ is commonly chosen through model-comparison methods: the user proposes a set of candidate λs and chooses one based on, e.g. BIC or cross-validation indices.

Of these common methods, using cross-validation to choose λ has been associated with overfitting, leading to an excess of false positives and low VSP (James and Radchenko, 2009; Meinshausen and Bühlmann, 2010). Fan et al. (2012) attribute this to the large sample correlations that can arise between noise predictors and the outcome. They demonstrate empirically that noise correlations increase in magnitude with increasing numbers of noise predictors. Further, when sample size N is less than number of predictors p, the largest noise correlation can easily exceed true correlations in magnitude. In cross-validation, the sample size used for model-fitting is always smaller than that in the entire sample, exacerbating the problems identified by Fan et al. (2012).

There has been little investigation into using bootstrap SEs and CIs to mitigate low VSP associated with cross-validated selection of λ. Asymptotic analyses have found that under bootstrap resampling, Lasso coefficients for noise predictors are expected to fluctuate between positive and negative values (Chatterjee, 2011; Camponovo, 2014). This should lead noise predictors to have larger bootstrap SEs and CIs than important predictors, which suggests that these statistics are useful for identifying false positive associations.

To investigate this expectation, we introduced and evaluated a bootstrap-based method with the goal of improving Lasso VSP by excluding false positive predictor selections under cross-validation. We investigated bootstrap SEs and CIs for the Lasso when cross-validated selection of λ is done before bootstrapping, which is the standard approach (e.g. D'Angelo et al. 2009; Sartori 2009). An alternative is to nest cross-validated selection of λ within each bootstrap replication, which could lead to larger SEs and wider CIs than in the standard approach (Buckland et al., 1997; Bühlmann et al., 2011). Such larger SEs and wider CIs lead to more conservative variable selection, and possibly to improved VSP. We compared the VSP resulting from the standard approach to bootstrapping to the VSP resulting from nested selection of λ.

We did this comparison in simulated and empirical data. The simulated data were generated to resemble data observed in genetic association studies. The simulated data were high-dimensional, with a small number of weak predictors and a large number of noise predictors. The empirical data were drawn from a genome-wide association study (GWAS). We made this choice because of the prominence of genetics, in particular GWAS and the use of polygenic (risk) scores, as a context for the application and development of the Lasso (Waldron et al., 2011; Lange et al., 2014).

2 Approach

Estimation of SEs and CIs for nonzero Lasso coefficient estimates may provide a way to control the Lasso's VSP as well as to assess the relative importance of predictors. In general, Lasso SEs or CIs cannot be estimated using a closed form (Osborne et al., 2000). A large variety of approaches has been tried to estimate Lasso SEs and CIs: see review paragraphs in e.g. Bühlmann et al. (2014), Chatterjee (2011), Kyung et al. (2010). Bootstrap estimation has received substantial interest, but many issues remain less explored, most importantly, the effects of choosing λ through cross-validation when using the bootstrap.

Next, we briefly review the two most common approaches to bootstrapping Lasso SEs and CIs: vector bootstrapping and residual bootstrapping. We follow this with a selective review of applied and methodological research using these approaches. We review both approaches to show that the behavior of the residual bootstrap has been studied in detail, but that comparatively little methodological research has been done on the vector bootstrap, despite its popularity in applied research. Hence, this paper focuses on the vector bootstrap.

2.1 Vector and residual bootstrapping

In nonparametric bootstrapping, samples are repeatedly drawn from observed sample data. The statistic of interest is calculated in each bootstrap sample. In this case, Lasso coefficient estimates are calculated, leading to an approximate sampling distribution. The approximate sampling distribution of Lasso coefficient estimates then permits calculation of SEs and CIs (Efron and Tibshirani, 1994).

We denote an estimate of the sampling distribution for the Lasso coefficient for predictor j as $\hat{F}(\hat{\beta}_j)$. $\hat{F}(\hat{\beta}_j)$ represents the marginal distribution of $\hat{\beta}_j$ values, as estimated using the nonparametric bootstrap. Two ways to use nonparametric bootstrapping to find $\hat{F}(\hat{\beta}_j)$ are vector bootstrapping and residual bootstrapping (Sartori, 2009).

2.1.1 The vector bootstrap

Vector bootstrapping begins with an observed sample of N observations measured on p predictors $x_1, \ldots, x_p$ and an outcome y. Each observation in the sample is considered as a row vector $\mathbf{z}_i=(x_{i1}, \ldots, x_{ip}, y_i)$, which consists of p predictor values x_{ij} and a single outcome value y_i. A Lasso model can be fit to every bootstrap sample of N observations $\mathbf{z}_i$, yielding a set of Lasso coefficient estimates (Camponovo, 2014). $\hat{F}(\hat{\beta}_j)$ is defined as the distribution of $\hat{\beta}_j$ values from each of all possible bootstrap samples of size N. The number of unique bootstrap samples increases faster than exponentially in N; to save computation time in practice, $\hat{F}(\hat{\beta}_j)$ is estimated using Monte Carlo simulation. The Monte Carlo estimate of $\hat{F}(\hat{\beta}_j)$ is written $\hat{F}^*(\hat{\beta}_j)$.

2.1.2 The residual bootstrap

Residual bootstrapping begins by fitting a linear regression model to a sample of N observations measured on p predictors $x_1, \ldots, x_p$ with outcome y, and generating the N residuals e. Residual bootstrapping uses the sampling distribution of residuals to simulate the distribution of y values about their conditional means $\mathbf{X}\beta$. Importantly, this requires treating the observed predictor values $x_1, \ldots, x_p$ as fixed and assuming that only the correct predictors are in the model (Efron and Gong, 1983), likely an inappropriate assumption in the context of variable selection. The residuals are then resampled.

Each bootstrap sample of N residuals, stored in the vector e^*, can be used to generate N outcomes y^*, defined as $y^*=\mathbf{X}\hat{\beta}+e^*$. The y^* values are regressed on $\mathbf{X}$ using the Lasso, yielding, as in vector bootstrapping, a set of Lasso coefficient estimates for each resample. This set of coefficient estimates is used to define $\hat{F}(\hat{\beta}_j)$, which is typically approximated through Monte Carlo simulation, as $\hat{F}^*(\hat{\beta}_j)$.

2.2 Previous research in bootstrapping the Lasso

2.2.1 Research with residual bootstrap

Residual bootstrapping of Lasso SEs and CIs has received more methodological research interest than has vector bootstrapping. Detailed investigations of the residual bootstrapped Lasso have been undertaken by Chatterjee (2011), Minnier et al. (2011), Kyung et al. (2010), and Knight and Fu (2000), among others, with the theoretical results in Chatterjee (2011) synthesizing much of the previous work.

By comparison, the behavior of Lasso SEs and CIs under vector bootstrapping has been under-studied, particularly with λ selected through cross-validation.

2.2.2 Research with vector bootstrap

The vector-bootstrapped Lasso has often been applied in statistical genetics, sometimes with λ selected through cross-validation. Further, the vector-bootstrapped Lasso is closely related to several other prominent variable-selection methods proposed in statistical genetics (Cho et al., 2010; Motyer et al., 2011; Valdar et al., 2012).

D'Angelo et al. (2009) proposed using vector resampling to estimate SEs of Lasso coefficients of SNP-SNP and gene-gene interaction terms. Sartori (2009) compared residual and vector bootstrapping of Lasso CIs and SEs in the context of statistical genetics, including the selection of λ through cross-validation before resampling. She found that: 1) residual and vector bootstrap SEs of Lasso coefficient estimates had similar degrees of bias in linear models; and 2) vector bootstrap CIs had superior coverage in linear and in logistic models. Camponovo (2014) used vector bootstrapping to generate simultaneous confidence regions in linear models with random predictors. He found poor coverage rates, and used an asymptotic argument to propose two modified vector bootstrapping procedures. The modified procedures generated confidence regions with adequate coverage.

The present study builds on the above research: in it, we compare the vector bootstrap with and without cross-validation nested within bootstrap replications. An important difference in the present study is that VSP, rather than coverage rate, is the criterion of comparison: if the CI for an important predictor excludes the true coefficient value but also excludes 0, the variable is correctly selected. Both versions of the bootstrapped Lasso are straightforward to implement in existing statistical software; e.g. `R` package `glmnet`, PLINK 1.9. (Friedman et al., 2010; Chang et al., 2014). This makes them attractive and approachable to applied researchers, who would benefit from understanding the trade-off in VSP involved in choosing one over the other.

2.2.3 Research with cross-validated selection of λ

Many methods have been proposed to increase the VSP of Lasso regression and related methods (Chatterjee, 2011; Fan et al., 2012; Lockhart et al., 2013). However, little methodological research has addressed the combination of vector bootstrapping and cross-validated selection of λ that is used in practice (Sartori, 2009; Cho et al., 2010). Despite the suggestion, in a recent textbook, that cross-validated selection of λ should be nested within bootstrap samples when applying the Lasso (Bühlmann et al., 2011), little published work has evaluated this procedure (Okser et al., 2014) The goal of the present paper is to address this deficiency and to investigate the conditions in which nested selection of λ leads to improved VSP. In addition, we address the question whether bootstrapped *t*-statistics are useful for the identification of false positives.

2.3 Role of λ when fitting Lasso models

The metaparameter λ controls the bias and parsimony of a fitted Lasso model. Equation 1 gives the definition of a Lasso model for predictors $\mathbf{X}$ and outcome y when λ is known (Tibshirani, 1996).

$$\hat{\beta}=\arg_{\beta}\min\frac{1}{2}\sum_{i}(y_i-\sum_{j}x_{ij}\beta_j)^2+\lambda\sum_{j}|\beta_j| \tag{1}$$

The value of λ determines the degree of shrinkage toward 0 and serves as a threshold for variable selection. A predictor is selected if the absolute value of its covariance with the outcome is larger than λ, and excluded otherwise (Efron et al., 2004). This thresholding property limits the useful range of values that λ can take. The minimum value that λ can take is 0, where the Lasso fit is the same as that of OLS regression. Such solutions are unbiased, but, because of the improbability of any OLS coefficients equaling 0 exactly, they are also unparsimonious.

The maximum value that λ can take depends on the largest sample covariance of any predictor with the outcome. More specifically, when λ is equal to or greater than that covariance, all coefficients are shrunk to 0, and the fitted model is intercept-only, thus parsimonious but biased (Friedman et al., 2007).

2.3.1 Selection of λ through *K*-fold cross-validation

In general, each value of λ is associated with a single Lasso model (Efron et al., 2004; Tibshirani, 2013).

K-fold cross-validation (Zhang, 1993) is often used to select the best-performing model. Lasso model fitting, λ selection, and *K*-fold cross-validation has been described in detail for its implementation in the `R` package `glmnet` (Friedman et al., 2010).

Following this procedure, the λ value that is associated with the best-performing model is selected. The best-performing model is the one having the minimum cross-validation index, which is computed as the sum of squared residuals averaged over the *K* cross-validations. The selected λ value is then used to fit a finalized model by solving Equation 1 in the entire sample. This produces a set of selected predictors that are then indexed in set **s**.

Different λ values might be selected in different samples from the same population due to the influence of noise correlations (Fan et al., 2012). In the next section, we interpret selection of λ as a source of variation in Lasso coefficient estimates.

2.3.2 Lasso variance estimates: contribution of λ selection

The variance of Lasso coefficient estimates depends on the joint distribution of the p predictor variables $\mathbf{X}$ and the outcome y, as well as on the value of λ (Pötscher and Leeb, 2009). The conditional distribution of estimates for a single predictor, denoted $g_j(\hat{\beta}|\lambda)$, is the distribution of $\hat{\beta}_j$ coefficients at a fixed λ value. The marginal distribution, $h_j(\hat{\beta})$, is the distribution of $\hat{\beta}_j$ averaged over λ values.

The variance of $\hat{\beta}_j$ can be found using either g_j or h_j. Heuristically, h_j treats the selected λ value as a realization of a random variable (Zhang, 1993; Bühlmann et al., 2014). We argue that using h_j might improve VSP because using g_j treats λ as fixed, which can underestimate the variance of coefficients.

To support this claim, consider the inequality:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_j)&=E_{\lambda}\{\mathrm{Var}_{\beta}(\hat{\beta}_j|\lambda)\}+\mathrm{Var}_{\lambda}\{E_{\beta}(\hat{\beta}_j|\lambda)\}\\ &\geq\mathrm{Var}_{\beta}(\hat{\beta}_j|\lambda)\end{aligned} \tag{2}$$

(Chatfield, 1995). If $\hat{\beta}_j$ and λ were independent, then $\mathrm{Var}(\hat{\beta}_j)=\mathrm{Var}_{\beta}(\hat{\beta}_j|\lambda)$, and there would be little difference between the fixed and random λ approaches in practice. However, $\hat{\beta}_j$ and λ are not necessarily independent: the range of possible λ values is bounded by $(0, r_{\max})$. Thus, although using g_j (treating $\hat{\beta}_j$ and λ as independent) has the practical advantage of using fewer computational resources, it will only be acceptable if the resulting underestimate of the standard error of $\hat{\beta}_j$ is small.

2.3.3 λ selection in bootstrapping

In practice, both g_j, the conditional distribution of $\hat{\beta}_j$ given λ, and h_j, the distribution of $\hat{\beta}_j$ averaged over all λs, are unknown. Both distributions can be estimated using the vector bootstrap.

Finding the bootstrap estimate of the conditional distribution, $\hat{g}_j^*$ is done by fitting Lasso models to resampled *X* and *y* values, given the λ value chosen through *K*-fold cross-validation in the original sample.

Finding the bootstrap estimate of the marginal distribution $\hat{h}_j^*$ requires treating the selected λ value as random. Nesting λ-selection within each bootstrap replication approximates the effect of sampling error on the value of λ selected.

Our simulations compared the fixed- and random-λ approaches with respect to VSP, and suggest effect sizes at which the increased computational burden of the random approach is worthwhile.

3 Methods

The purpose of the current paper is to propose and to evaluate the use of the vector bootstrap, with λ selected through *K*-fold cross-validation, as a method for estimating Lasso SEs and CIs. In particular, we compared three variants of this approach: a software default approach to variable selection using the Lasso (Method 1, see Figure 1); an approach involving selection of λ before resampling (Method 2, see Figure 2); and a third approach where λ-selection is nested within bootstrap samples (Method 3, see Figure 3). Our evaluation was in terms of VSP and of accuracy of ranking predictors by relative importance. Relative importance was calculated using coefficients of variation, ($|t^{-1}|$, where *t* is a bootstrapped *t*-statistic).

In the first step of each variable selection method, a Lasso model is fit to the entire sample. This requires selection of λ, which is done through *K*-fold cross-validation. The initial model fit produces a set of selected

```
Given $N$ individuals measured on $p$ standardized predictors $x_j$, $j = 1, \ldots, p$, with outcome $y$:
1.  Fit Lasso model to select predictors
    (a) Identify candidate $\lambda$ values using sample covariances
    (b) Use $K$-fold cross-validation ($K$=10) to select finalized $\lambda$
2.  Finalized $\lambda$ selects predictors, indexes them in $\mathbf{s} \subset \{1, \ldots, j\}$
3.  Exclude predictors $x_k$, where $k \notin \mathbf{s}$
```

Figure 1: Method 1–default approach to Lasso variable selection using *K*-fold cross-validation.

```
1.  Fit Lasso models, select set of predictors $\mathbf{s}$ as in Method 1
2.  Estimate marginal sampling distributions $\hat{F}^*(\hat{\beta}_k)$ for $k \in \mathbf{s}$
    (a) Draw $B$ vector bootstrap samples ($B \geq 1000$)
    (b) In each bootstrap sample:
        i.  Fit a Lasso model with $\lambda$ at finalized value from Step 1 (Fixed $\lambda$)
3.  Use $\hat{F}^*(\hat{\beta}_k)$ to calculate mean, SE, and CI for each $\hat{\beta}_k$
4.  Exclude $x_k$s:
    if  CI for $\hat{\beta}_k$ includes 0
            or
    if  $C^*_{\text{var}} = (t^*_k)^{-1} = \frac{\text{SE}^*(\hat{\beta}_k)}{|\text{Mean}^*(\hat{\beta}_k)|} > 0.5$ (or other cutoff)
```

Figure 2: Method 2–vector bootstrap for improved Lasso variable selection precision with λ treated as fixed.

```
1. Fit Lasso models, select set of predictors s as in Method 1
2. Estimate marginal sampling distributions $\hat{F}^*(\hat{\beta}_k)$ for $k \in \mathbf{s}$
    (a) Draw $B$ vector bootstrap samples ($B \geq 1000$)
    (b) In each bootstrap sample:
        i. Fit a Lasso model exactly as in Step 1 (Random $\lambda$)
3. Use $\hat{F}^*(\hat{\beta}_k)$ to calculate mean, SE, and CI for each $\hat{\beta}_k$
4. Exclude $x_k$s using rules listed in Method 2
```

Figure 3: Method 3–vector bootstrapping with metaparameter λ treated as random.

predictors, which are indexed in the set **s**. This step is the default application of the Lasso in the `R` packages `glmnet` and `grpreg`. We denote it Method 1.

Methods 2 and 3 differ from Method 1 by having a second variable selection step. In this step, further reduction of the set of selected predictors is done using bootstrap SEs or CIs. All predictors are used in vector bootstrap resampling, but, to save computational resources, SEs and CIs are not calculated for predictors that were excluded in the initial variable selection step.

Method 2 uses the same value of λ in every bootstrap sample.

Method 3 differs from Method 2 by re-selecting λ in each bootstrap sample. In both methods, after SEs or CIs are calculated, variables that: 1) include 0 in their confidence intervals; or that: 2) have a coefficient of variation greater than a certain cutoff; are excluded.

3.1 Lasso CIs and SEs: secondary selection or ranking

Bootstrap CI or SE estimates improve Lasso models through a second step of selecting or ranking predictors. CI and SE estimates are both directly related to the sampling variance of a Lasso coefficient estimate, discussed above. A 1–α CI for the coefficient estimate $\hat{\beta}_j$ is generated either using the $\frac{\alpha}{2}$, $1-\frac{\alpha}{2}$ quantiles of the bootstrap distribution $\hat{g}_j^*$, or using an approximate inverted z-test, which gives the interval $\hat{\beta}_j \pm z_{\alpha/2}\mathrm{SE}^*(\hat{\beta}_j)$, where $z_{\alpha/2}$ is the $\frac{\alpha}{2}$ quantile of a standard normal distribution and $\mathrm{SE}^*(\hat{\beta}_j)$ is the bootstrap estimate of the standard error of $\hat{\beta}_j$.

Using either CI method, predictors that have CIs that contain 0 are excluded since this can be regarded as evidence that predictor x_j is a false positive selection.

SEs can also be used to give information about the relative importance of predictors in addition to improved VSP. We propose the use of the coefficient of variation for each nonzero Lasso regression weight as an index of relative importance and to apply cutoffs to this statistic to exclude false positives. The coefficient of variation of a random variable X, denoted $C_{\mathrm{var}}(X)$, is the ratio of its standard error to the absolute value of its mean.

This index is sensitive to small differences in mean values and thus may be better able to distinguish small true positives from false positives. We used the vector bootstrap to estimate C_{var} for individual Lasso coefficients, denoted $C^*_{\mathrm{var}}(\hat{\beta}_j)$.

4 Simulation studies

We first compared Methods 1–3 using a factorial simulation study. The simulation had two goals: first, evaluating the Methods' ability to distinguish signal from noise; second, evaluating their ability to correctly order signals of differing strengths.

The data generation models were as simple as possible while still representing two empirically interesting scenarios based on statistical genetics: 1) a low probability of selecting important predictors at random; and 2) a spectrum of small true effect sizes. To this end, data were generated under two different linear models: first, a few-important-predictors model with under 5% of predictors having true effects, and with effect sizes (given in R^2) representing 2.5% or less of outcome variance attributable to any important predictor; and second, a polygenic model, in which there were thousands of predictors, each of which had an effect drawn from a normal distribution, all of which together accounted for 60% of the variance in the outcome.

Two thousand five hundred Monte Carlo (MC) replications were used in each cell of the few-important-predictors design. This number of replications was chosen based on pilot studies, in which at least 2500 replications were required in order to generate relatively smooth empirical distributions of coefficient estimates (see also Sartori 2009). However, only 250 MC replications were used when data were simulated under the polygenic model due to the substantial use of computational resources needed to bootstrap such data.

B=1000 bootstrap replications were used within each MC replication. The average performances of the three methods across samples were compared; within each MC replication, each method was employed on an independent sample drawn from the population distribution. This was done in order to avoid creating dependence among results that might arise from fitting the methods to the same data.

4.1 Simulation design

Three factors were manipulated in the simulations: method, data-generating model, and effect size.

As described above, the methods compared were the fixed- (Method 2) and random-λ (Method 3) variants of the vector bootstrapped Lasso, with the default application, Method 1.

The second factor manipulated in the simulation study was the data-generating model. Three data-generating models were used: a single important predictor and 99 noise predictors; a five important predictors and 99 noise predictors, with the important predictors having different R^2 values, enabling us to rank them; and a polygenic model with 3000 predictors with effect sizes drawn from a $\mathcal{N}(0, 0.60)$ distribution–the three predictors with the largest (absolute) effects were treated as the important predictors.

Each data generating model was a linear regression model having a standard normal outcome and binomial (2, 0.5) distributed predictors; N=2500 was used as the sample size. This was chosen as a rough approximation of the sample size and predictor structure of smaller genome-wide association studies of quantitative phenotypes (Balding, 2006).

The third factor manipulated in the simulation study was effect size. Effect sizes of r^2=0.01, 0.0033, 0.001 were used in the single-important-predictor analyses. In the five-important-predictors analyses, each important predictor had a different effect size: the set R^2=0.01, 0.0067, 0.0033, 0.0022, 0.001 was used. In the polygenic model, the three strongest predictors were expected to account for 5.1% of phenotypic variance together. This estimate is based on treating the largest simulated-SNP effects as being drawn from a truncated normal distribution representing the upper 0.15% tail area of the $\mathcal{N}(0, 0.60)$, and then treating the strongest negative effects as being drawn independently from the corresponding part of the lower tail (Barr and Sherrill, 1999).

We manipulated the effect size of important predictors for two reasons: we used effect size as a measure of the “difficulty” of correctly selecting important predictors, giving us a way to use the data to influence the methods’ VSP; and because previous simulation studies (e.g. Leng et al., 2006; Meinshausen and Bühlmann, 2010; Tibshirani, 2011), used effect sizes that are now considered to be unrealistically large in the context of statistical genetics (Stefansson et al., 2009; Park et al., 2011).

In the next section, we describe the evaluation criteria.

4.2 Evaluation criteria

The main question asked in this paper is: when does nested selection of λ lead to improved VSP over other approaches? A subsidiary question is: can C^{*}_{var}-statistics give information about the relative importance of predictors that might be used to identify false positives? Addressing these questions requires quantifying the performance of the different methods. We used VSP to quantify the methods' performance and additionally used the False Negative Rate (FNR) to identify the risk that each method might be over-conservative.

$$\text{VSP is } \frac{\#\text{Important predictors selected}}{\#\text{Selected predictors}} = \frac{\#\text{True positives}}{\#\text{Positives}}.$$

VSP is set to 0 if there are no positives. Thus, in each replication of the few-important-predictors simulation, VSP ranged from $0, \frac{1}{104}, \frac{1}{103}, \dots, \frac{1}{2}, 1,$ while in the polygenic model, the denominator was 3000. FNR is the proportion of truly important predictors that have been classified as unimportant by a variable selection method (Fawcett, 2006).

A positive in the vector bootstrap Lasso (Methods 2 and 3) was defined as a predictor for which the (1–α)×100% bootstrap percentile confidence interval (Efron and Tibshirani, 1994) excluded 0. Confidence level (α) was set to 0.05, and intervals were symmetric. In sensitivity analyses, confidence levels of 0.02, 0.10 and 0.20 were also used.

A positive in the default Lasso (Method 1) was defined as a predictor having a coefficient in the finalized Lasso model (i.e. a predictor indexed in **s**).

To quantify the performance of C^*_{var} as an importance measure, we used the median rank and median absolute deviation of ranks of each predictor's C^*_{var} over MC replications. We chose this measure to obtain both a typical rank and the variability of rankings for important and for noise predictors. This importance measure was not used with data that had been simulated under a polygenic model because the randomness of the effects meant that there was no consistent mapping between a predictor's index and its effect size.

4.3 Simulation results

4.3.1 Improved variable selection precision

Use of the vector bootstrap (Methods 2 and 3) was associated with increased VSP at all effect sizes, but also with increased FNR at all effect sizes. This is shown in Table 1, for nominal α=0.05 and percentile bootstrap confidence intervals; normal-theory bootstrap confidence intervals are not shown because their performance was similar to, but slightly worse than that of the percentile intervals. At the smaller effect sizes, the increased precision was only apparent with random λ, and, at the smallest effect size, the magnitude of this advantage was small. The increased FNR suggests that this pattern was due to the bootstrapping procedures being more conservative than Method 1.

Table 1: Variable selection precisions and false negative rates of default Lasso vs. bootstrapped percentile CI.

	R^2\Method	Default	Fixed λ	Random λ
VSP	Ranks	0.3646	**0.9529**	0.8956
	0.01	0.3720	0.8304	**0.8528**
	0.0033	0.2049	0.1536	**0.3372**
	0.001	0.0503	0.0096	**0.0507**
	Polygenic	0.0096	**0.1154**	0.0990
FNR	Ranks	**0.2008**	0.5784	0.4594
	0.01	**0.0084**	0.1500	0.0408
	0.0033	**0.3776**	0.8420	0.6088
	0.001	**0.7928**	0.9888	0.9404
	Polygenic	**0.1343**	0.8804	0.8372

Bold text indicates the largest VSP/smallest FNR in each row. Nominal α=0.05.
The random-λ bootstrap CI (Method 3) is more precise than the default at each effect size, but improvement is marginal for the very smallest effects. In all cases, the default approach showed low false-negative rates.

In sensitivity analyses, we found that the nominal coverage rate α chosen for confidence intervals interacted with the bootstrapping method used. For example, at R^2=0.0033, the random approach (Method 3) is more precise than the fixed approach (Method 2) at α=0.01, 0.05 (Table 1), but less precise at α=0.10 (not shown).

The bootstrapped coefficient of variation (C^*_{var}) was in general much larger for false positives than for true positives. Bootstrap means for noise predictors were never 0, preventing the occurrence of division-by-0 errors in computing C^*_{var} values. These observations support the use of a C^*_{var} cutoff as a way of increasing Lasso VSP. However, C^*_{var} cutoffs did not increase VSP to the extent that CIs did. They offered no improvement over the default Method 1 at effect sizes R^2=0.0033, 0.01 or in multiple-predictor models.

4.3.2 Ranking predictors

The usefulness of C^*_{var} to rank predictors by relative importance depended on the effect size as well as the number of true predictors. At R^2=0.0033, 0.01 or with multiple true predictors, using C^*_{var} to rank predictors by relative importance yielded no improvement over ranking predictors by the absolute value of their non-bootstrapped coefficients. This is shown in Table 2. Additionally, true predictor ranks were identical regardless of whether λ was treated as random or as fixed.

For a single true predictor with R^2=0.001, ranking by C^*_{var} led to better discrimination of the true predictor than did ranking by non-bootstrapped coefficients. The true predictor did not always have the smallest C^*_{var}, but that it was within the top six predictors at least half the time. Without bootstrapping, the true predictor was often selected out of the model. Overall, it was in the top 50% of predictors close to half the time but that it was not frequently among the highest ranks.

4.3.3 Variability of coefficients under different methods

Treating the metapararameter λ as random was associated with greater variability of coefficient estimates, as illustrated in Table 3 by larger standard deviations, and wider confidence intervals for Method 3 when

Table 2: Median and MAD of ranks of five important predictors out of 104 total.

R^2\Method	Default (1)	Bootstrapped (C^*_{var}) (2 and 3)
0.01	1 (0)	1 (0)
0.0067	2 (0)	1 (0)
0.0033	3 (0)	3 (0)
0.0022	4 (1.48)	4 (1.48)
0.001	5 (2.97)	6 (4.48)

Bootstrapping resulted in no improvement of predictor ranks. Both bootstrapping methods (2 and 3) showed identical performance.

Table 3: Standard deviations and confidence interval lengths of Lasso estimates of important predictors.

Model	Method	SD	qCI Length	ntCI Length
Ranks	2 (Fixed λ)	0.0099	0.0351	0.0405
	3 (Nested λ)	0.0177	0.0693	0.0692
Single/R^2=0.0033	2 (Fixed λ)	0.00637	0.02060	0.02500
	3 (Nested λ)	0.0167	0.0650	0.0656
polygenic	2 (Fixed λ)	0.0172	0.0604	0.0674
	3 (Nested λ)	0.0239	0.0860	0.0857

SD, Standard deviation; qCI, bootstrap quantile confidence interval; ntCI, bootstrapped normal approximation confidence interval; nominal α=0.05.
Measures of the variability of coefficient estimates for important predictors, averaged over MC replications, showed that Method 3 had increased variability compared to Method 2.

compared to Method 2. Table 3 presents summaries for important predictors only; results for noise predictors were very similar (results in the table are averages over MC replications and are not conditional on the important predictors being selected into the model). Confidence interval lengths were similar between quantile-based bootstrap CIs and normal-theory bootstrap CIs.

These results are consistent with inequality (2), which suggests that the increased variability is attributable to the variance in expected coefficient values with respect to the distribution of λ. Table 4 compares five-number summaries of the distribution of λ from Method 1 to those from Method 3. The results for Method 1 are five-number summaries of the λ values that were selected across replications, while those for Method 3 are the averages across replications of the five-number-summary of λ values. Selection of λ via Method 3 (cross-validation nested within bootstrap replications) tends to produce lower values of λ (distribution shifted left) which are also less-variable (smaller IQR). The smaller median λ means that Method 3 performed less regularization than did Method 1, hence had coefficients with larger values and may have included more predictors in each bootstrap replication. Thus, bootstrap CIs from Method 3 would have been relatively wide, leading to increased VSP because of liberal variable selection within each bootstrap replication.

The results of the simulation studies quantified the methods' relative performance in idealized data. For a more critical evaluation of their practical utility, we applied them in a GWAS data set, using them to select pairwise interactions as well as main effects.

5 Empirical illustration

We used data gathered for a genome-wide association study (GWAS) of Borderline Personality Disorder features to compare the vector bootstrap with λ-selection nested within bootstrap samples to the standard Lasso. This comparison serves as a representative analysis for possible applications of Method 3. The original GWAS was based on a sample of N=7124 individuals who participated in a twin-family study of mental and somatic health (Boomsma et al., 2006; Willemsen et al., 2013); see Willemsen et al. for detailed methods including IRB approval, genotyping, and quality control procedures. Responses to a psychiatric inventory measuring Borderline Personality features were used as outcomes because they have shown a promising signal that was replicated in an independent sample (Lubke et al., 2014).

Borderline features were measured using total scores on the PAI-BOR inventory, a 24-item test (Morey, 1991). More specifically, the outcome we used in this study was the residual of PAI-BOR score after OLS regression on age, gender, and their interaction, as well as a principal component score representing ancestry (Price et al., 2010; Abdellaoui et al., 2013). Following up on D'Angelo et al. (2009)'s proposal, we fit Lasso multiple regressions of Borderline features on SNP main effects and SNP-SNP interactions. To control the computational resources required, we limited the analysis to pairwise interactions and main effects of the 125 SNPs having the strongest univariate association with the Borderline features phenotype, as listed in Lubke et al. 2014. R's memory limitations limit the application of bootstrapping to interactions between 1500 or

Table 4: Five-number summaries of the distribution of λ values in different simulations.

Model	Method	Min	1Q	Med	3Q	Max
Ranks	1 (CV only)	0.0154	0.0282	0.0324	0.0367	0.0582
	3 (Nested λ)	0.00753	0.01410	0.01640	0.01910	0.04030
Single/R^2=0.0033	1 (CV only)	0.0196	0.0365	0.0446	0.0505	0.0685
	3 (Nested λ)	0.00812	0.01540	0.01820	0.02190	0.06020
Polygenic	1 (CV only)	0.00796	0.01320	0.01490	0.01680	0.03180
	3 (Nested λ)	0.00182	0.00463	0.00543	0.00625	0.00968

Summaries of the distribution of λ values suggest that Method 3 (selection nested within bootstrap samples) produces a less-variable, left-shifted distribution.

fewer variables ($1500^2 \times 1000$ bootstrap samples $\approx 2^{31}$ objects) (R Core Team, 2013). To avoid multicollinearity, these 125 were then pruned to the set of 77 SNPs that had pairwise correlations of less than r=0.6 among each other. From these, 2926 pairwise interaction terms were calculated, resulting in a total p=3003, and hence up to 3003 Lasso partial regression weights needing CIs and C^*_{var}.

The primary purpose of this analysis was to compare the different methods for lasso variable selection in a data set with SNPs having different allele frequencies and unknown effects on the outcome. The secondary purpose was an exploratory analysis of epistatic effects between SNPs as predictors of borderline personality symptoms, generating hypotheses that can be tested in independent samples. We did not emphasize the need for the selected models to be biologically plausible; that is, we did not group SNPs for selection by gene or pathway. We did not rule out models of pure interaction in the absence of main effects (Cordell, 2009). Accordingly, we did not force selection of hierarchical models: that is,it was not the case that any interaction terms considered for selection must have had main effects in the model. Because of this and because of the bias caused by Lasso estimation, the values of the coefficients for (pure) interaction terms cannot be interpreted straightforwardly as magnitudes of effect modifications.

Results from applying the vector bootstrap with random λ (Method 3) were compared to those from the default Lasso (Method 1). Method 3 was used to generate percentile CIs and C^*_{var}. The resulting variable selections and rankings were compared to Method 1's selections and to ranking coefficient estimates from the finalized model by absolute value.

5.1 Empirical illustration: results

The empirical illustration concerned the application of the bootstrapped Lasso to pairwise interaction effects. The bootstrapped Lasso produced different variable selections and importance rankings than did the default Lasso. The bootstrapped Lasso tended to select better quality predictors than did the default Lasso. The default Lasso selected predictors with low minor allele frequency (MAF), hence had very large standard error estimates. This suggests that the default Lasso can ignore important aspects of the data.

Figure 4 and Table 3 present comparisons of the default Lasso and the approach using vector bootstrapping. Figure 4 plots bootstrap mean estimates on the horizontal axis and bootstrap standard error estimates on the vertical axis. Points falling outside of the dark gray **V** have C^*_{var} values less than 0.5 (bootstrapped t-statistics greater than 2). Predictors that were selected by the default Lasso are plotted as light gray diamonds. There is no obvious relationship between default Lasso selection of a predictor and its bootstrap

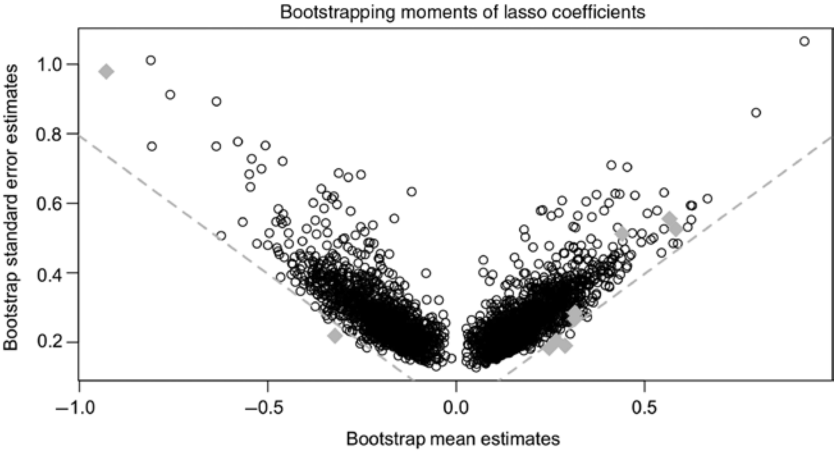

Figure 4: Bootstrap means and SEs of 77 SNPs, 2926 pairwise interactions; light gray diamonds represent predictors selected without bootstrapping. The lines represent mean $=\pm 2\times \hat{SE}^*$: points and diamonds ouside the **V**-shape (i.e. in lower corners) are promising signals. A cube-root transformation was used on both axes.

Table 5: Promising SNP-SNP interactions.

Chrs	rsIDs	MAFs	C^*_{var}	CI
16, 1	rs118160379×rs59194015	0.05, 0.27	0.49	(−0.078, −0.002)
Selected by default but rejected by bootstrap				
9, 4	rs112188788×rs139344595	0.02, 0.01	0.73	(0, 501)
6, 1	rs117666484×rs73008417	0.01, 0.01	0.94	(0, 562)
9, 1	rs112188788×rs73008417	0.02, 0.01	1.17	(−2.904, 0)
12, 9	rs117256451×rs112188788	0.02, 0.02	1.57	(−0.056, 451)

moments. dbSNP lookup of the predictors in Table 5 showed that the bootstrapped Lasso was less prone to selecting interactions between SNPs having low MAF than was the default Lasso. The default Lasso, in selecting these interactions, was in effect including interactions between binomial predictors that have low success probabilities. These interactions tended to have large bootstrap standard errors, hence are excluded when a C^*_{var} cutoff is used. Interestingly, the low-MAF SNPs involved these interactions tended to have moderately strong main effects in the conventional GWAS analyses.

Using vector bootstrapping of Lasso coefficients (CI or C^*_{var} cutoff of 0.5) suggested a single interaction for followup.

6 Conclusion

Using vector bootstrap CIs on Lasso regression coefficients offers a valid way to distinguish false positive selections from true positives. Percentile CIs were associated with increased precision of variable selection at all effect sizes. At the smallest effect sizes, including those in a polygenic model, gains were only achieved when using Method 3, which treated the metaparameter λ as random. Additionally, the bootstrapped methods were more conservative in variable selection than was Method 1, the default Lasso. This suggests that, if several small effects are expected and if avoiding false positives is more important than avoiding false negatives, treating λ as random justifies increased computational cost.

The (bootstrapped) coefficient of variation, C^*_{var} does measure the relative importance of Lasso predictors, but offers little to recommend it over using the absolute value of Lasso coefficients.

We observed a “rising tide lifts all boats” effect for all methods, where a predictor having a given small effect size was more likely to be selected when the data were generated to have other important predictors. This was despite the predictors and their effects being independent of one another. The more complex models tended to have lower (more lenient) thresholds λ selected by cross-validation, regardless of the method used. A possible explanation is that a λ causing inclusion of a solitary small effect might not be able to consistently decrease the residual sum of squares in different cross-validation subsamples, but that a λ that admits multiple small effects could.

The low degree of overlap in the distributions of λ estimated by Methods 1 and 3 suggests a need for explanation. The two distributions should have the same mean value, because the bootstrap distribution of a statistic should approximate the distribution of a statistic across repeated independent samples.

In the empirical analysis, the vector-bootstrapped Lasso excluded unreliable predictors that had been selected by the default Lasso. However, it is possible that residual bootstrapping could have led to better performance with low MAF predictors. Under vector bootstrapping, “monoallelic” SNPs are possible within the bootstrap samples. The result would be inflation of the intercept term in the regression model, which would affecting model-fitting through cross-validation. Further, VSP is dependent on the base rates of positives and negatives, and would have been skewed if important SNPs tended to have low MAFs and noise SNPs high MAFs, or vice-versa, which limits the generalizability of the simulation results to empirical data.

Three follow-up studies are suggested by this result: a simulation study comparing the two Lasso methods after manipulating the reliability of predictors; an attempt to replicate the promising interaction between rs118160379 and rs59194015; and a comparison of VSP from Method 3 to that from other Lasso confidence intervals, e.g. those in Camponovo (2014).

In addition, both our empirical and our simulation results are also relevant to research that uses polygenic scores to predict complex traits and to investigate the polygenic architecture of closely related traits. Lasso regression has recently been implemented in PLINK 1.9, a software package that is very widely used in the analysis of complex traits. In consequence, Lasso variable selections are being used to construct polygenic scores, but have not yielded significant improvements over conventional methods (Warren et al., 2014).

Our simulation results suggest that vector bootstrapping (with nested selection of λ) may be able to yield polygenic scores with greater predictive accuracy, but will require relatively large sample sizes to avoid excluding important variants from the score. Large samples are relatively common in the GWAS context, however using this approach efficiently will require careful application and data management.

There were several limitations to this study. The data generating model had predictors that were independently and identically distributed as well as a normally distributed outcome. These attributes are unlikely to hold in empirical data. We plan to extend the current simulations with correlated and differently scaled predictors as well as skewed outcomes. Second, the simulated effect sizes, while small, were still somewhat larger than those that are typically observed in the statistical genetics of complex traits (Stefansson et al., 2009). The phenotypic variance explained by the polygenic model was consistent with a highly heritable trait, e.g. height, and is larger than would that expected for most complex traits of interest. The number of important predictors used was also much smaller than the number of genetic loci expected to influence complex phenotypes (Sivakumaran et al., 2011). Similarly, our results suggest that bootstrapping is most useful when there are many predictors to consider, to be reduced to a relatively small number of important ones, and that the single-important predictor case (tested here) is perhaps suboptimal for evaluating the precision and conservatism of the bootstrap methods. Finally, the argument used to justify nesting λ-selection within bootstrapping was intuitive. A more rigorous argument might be able to identify specific conditions on $\mathbf{X}$ or y that would lead to Method 3 consistently outperforming Method 2, or vice-versa. On the other hand, simulations could be used to estimate the components of bootstrapped variance due to the individual terms in Equation (2).

Vector bootstrapping CIs of Lasso coefficients led to increased VSP, especially at small effect sizes. Our illustration with empirical data showed that this is also an effective approach to select important interactions between predictors. In consequence, vector bootstrapping CIs is a very promising approach for identifying sets of SNP-SNP and SNP-environment interactions.

Funding: Netherlands Organization for Scientific Research, (Grant/Award Number: '904-61-090', '985-10-002', '400-05-717', '904-61-193480-04-004', 'Addiction-31160008', 'Middelgroot-911-09-032', 'Spinozapremie 56-464-14192'), European Science Foundation, (Grant/Award Number: 'EU/QLRT-2001-01254'), European Research Council, (Grant/Award Number: 'ERC Advanced, 230374'), Biobanking and Biomolecular Resources Research Infrastructure, (Grant/Award Number: 'BBMRI –NL, 184.021.007'), National Institutes of Health, (Grant/Award Number: 'R01 HD042157-01A1', '1RC2 MH089951', 'MH081802'), Center for Medical Systems Biology (CSMB, NWO Genomics), [Grant/Award Number: 'NBIC/BioAssist/RK(2008.024)'], National Institute of Mental Health, (Grant/Award Number: 'NIMH U24 MH068457-06'), NSF-MRI, (Grant/Award Number: 'BCS-1229450'), NIDA, (Grant/Award Number: 'NIDA-DA018673').

References

Abdellaoui, A., J.-J. Hottenga, P. de Knijff, M. G. Nivard, X. Xiao, P. Scheet, A. Brooks, E. A. Ehli, Y. Hu, G. E. Davies, J. J. Hudziak, P. F. Sullivan, T. van Beijsterveldt, G. Willemsen, E. J. de Geus, B. W. Penninx and D. I. Boomsma (2013): "Population structure, migration, and diversifying selection in the netherlands," Eur. J. Hum. Genet., 21, 1277–1285.

Ayers, K. L. and H. J. Cordell (2010): "SNP Selection in genome-wide and candidate gene studies via penalized logistic regression," Genet. Epidemiol., 34, 879–891.

Balding, D. J. (2006): "A tutorial on statistical methods for population association studies," Nat. Rev. Genet., 7, 781–791.

Barr, D. R. and E. T. Sherrill (1999): "Mean and variance of truncated normal distributions," Am. Stat., 53, 357–361.

Boomsma, D. I., E. J. C. de Geus, J. M. Vink, J. H. Stubbe, M. A. Distel, J.-J. Hottenga, D. Posthuma, T. C. E. M. Van Beijsterveldt, J. J. Hudziak, M. Bartels and G. Willemsen (2006): "Netherlands twin register: from twins to twin families," Twin Res. Hum. Genet., 9, 849–857.

Buckland, S. T., K. P. Burnham and N. H. Augustin (1997): "Model selection: an integral part of inference," Biometrics, 53, 603–618.

Bühlmann, P. L., S. A. van de Geer and S. Van de Geer (2011): Statistics for high-dimensional data methods, theory and applications, Springer, Heidelberg.

Bühlmann, P., L. Meier and S. van de Geer (2014): "Discussion: 'a significance test for the lasso'," Ann. Statist., 42, 469–477.

Camponovo, L. (2014): "On the validity of the pairs bootstrap for lasso estimators," Social Science Research Network Working Paper Series.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell and J. J. Lee (2014): Second-generation plink: rising to the challenge of larger and richer datasets, arXiv preprint arXiv:1410.4803.

Chatfield, C. (1995): "Model uncertainty, data mining and statistical inference," J. R. Stat. Soc. Series A, 158, 419–466.

Chatterjee, A. (2011): "Bootstrapping lasso estimators," J. Am. Stat. Assoc., 106, 608–625.

Cho, S., K. Kim, Y. J. Kim, J.-K. Lee, Y. S. Cho, J.-Y. Lee, B.-G. Han, H. Kim, J. Ott and T. Park (2010): "Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis," Ann. Hum. Genet., 74, 416–428.

Cordell, H. J. (2009): "Detecting gene–gene interactions that underlie human diseases," Nat. Rev. Genet., 10(6), 392–404.

D'Angelo, G., D. C. Rao and C. C. Gu (2009): "Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies," BMC Proc., 3(Suppl. 7), S62.

Devlin, B., K. Roeder and L. Wasserman (2003): "Analysis of multilocus models of association," Genet. Epidemiol., 25, 36–47.

Efron, B. and G. Gong (1983): "A leisurely look at the bootstrap, the jackknife, and cross-validation," Am. Stat., 37, 36–48.

Efron, B., T. Hastie, L. Johnstone and R. Tibshirani (2004): "Least angle regression," Ann. Stat., 32, 407–499.

Efron, B. and R. J. Tibshirani, (1994): An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability), Chapman and Hall/CRC.

Fan, J., S. Guo and N. Hao (2012): "Variance estimation using refitted cross-validation in ultrahigh dimensional regression," J. R. Stat. Soc. Series B (Statistical Methodology), 74, 37–65.

Fawcett, T. (2006): "An introduction to roc analysis," Pattern Recogn. Lett., 27, 861–874.

Freedman, D. and D. Lane (1983): "A nonstochastic interpretation of reported significance levels," J. Bus. Econ. Stat., 1, 292–298.

Friedman, J., T. Hastie, H. Höfling and R. Tibshirani (2007): "Pathwise coordinate optimization," Ann. Appl. Stat., 1, 302–332.

Friedman, J., T. Hastie and R. Tibshirani (2010): "Regularization paths for generalized linear models via coordinate descent," J. Stat. Softw., 33, 1.

He, Q. and D.-Y. Y. Lin (2011): "A variable selection method for genome-wide association studies," Bioinformatics, 27, 1–8.

James, G. M. and P. Radchenko (2009): "A generalized dantzig selector with shrinkage tuning," Biometrika, 96, 323–337.

Knight, K. and W. Fu (2000): "Asymptotics for Lasso-Type estimators," Ann. Stat., 28, 1356–1378.

Kyung, M., J. Gill, M. Ghosh and G. Casella (2010): "Penalized regression, standard errors, and bayesian lassos," Bayesian Anal., 5, 369–412.

Lange, K., J. C. Papp, J. S. Sinsheimer and E. M. Sobel (2014): "Next-generation statistical genetics: modeling, penalization, and optimization in high-dimensional data," Annu. Rev. Stat. Appl., 1, 279–300.

Leng, C., Y. Lin and G. Wahba (2006): "A note on the lasso and related procedures in model selection," Stat. Sinica, 16, 1273.

Lockhart, R., J. Taylor, R. J. Tibshirani, and R. Tibshirani (2013): A significance test for the lasso, arXiv preprint arXiv:1301.7161.

Lubke, G., C. Laurin, N. Amin, J. Hottenga, G. Willemsen, G. van Grootheest, A. Abdellaoui, L. Karssen, B. Oostra, C. M. van Duijn, B. W. Penninx, D. I. Boomsma (2014): "Genome-wide analyses of borderline personality features," Mol. Psychiatry, 19, 923–929.

Meinshausen, N. and P. Bühlmann (2010): "Stability selection," J. R. Stat. Soc. Series B (Statistical Methodology), 72, 417–473.

Minnier, J., L. Tian, and T. Cai (2011): "A perturbation method for inference on regularized regression estimates," J. Am. Stat. Assoc., 106, 1371–1382.

Morey, L. C. (1991): Personality assessment inventory: professional manual, Psychological Assessment Resources, Odessa, FL.

Motyer, A., C. McKendry, S. Galbraith and S. Wilson (2011): "LASSO model selection with post-processing for a genome-wide association study data set," BMC Proc., 5(Suppl. 9), S24.

Okser, S., T. Pahikkala, A. Airola, T. Salakoski, S. Ripatti and T. Aittokallio (2014): "Regularized machine learning in the genetic prediction of complex traits," PLoS Genet., 10, e1004754.

Osborne, M. R., B. Presnell and B. A. Turlach (2000): "On the LASSO and its dual," J. Comp. Graph. Stat., 9, 319–337.

Park, J.-H. H., M. H. Gail, C. R. Weinberg, R. J. Carroll, C. C. Chung, Z. Wang, S. J. Chanock, J. F. Fraumeni and N. Chatterjee (2011): "Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants," Proc. Natl. Acad. Sci. USA., 108, 18026–18031.

Pötscher, B. M. and H. Leeb (2009): "On the distribution of penalized maximum likelihood estimators: the LASSO, SCAD, and thresholding," J Multivar. Anal., 100, 2065–2082.

Price, A. L., N. A. Zaitlen, D. Reich and N. Patterson (2010): “New approaches to population stratification in genome-wide association studies,” Nat. Rev. Genet., 11, 459–463.

R Core Team (2013): “R: a language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria.

Sartori, S. (2009): “Penalized Regression: bootstrap confidence intervals and variable selection for high dimensional data sets,” PhD thesis, Universitá Degli Studi di Milano.

Sivakumaran, S., F. Agakov, E. Theodoratou, J. G. Prendergast, L. Zgaga, T. Manolio, I. Rudan, P. McKeigue, J. F. Wilson and H. Campbell (2011): “Abundant pleiotropy in human complex diseases and traits,” Am. J. Hum. Genet., 89, 607–618.

Stefansson, H., R. A. Ophoff, S. Steinberg, O. A. Andreassen, S. Cichon, D. Rujescu, T. Werge, O. P. Pietiläinen, O. Mors, P. B. Mortensen, E. Sigurdsson, O. Gustafsson, M. Nyegaard, A. Tuulio-Henriksson, A. Ingason, T. Hansen, J. Suvisaari, J. Lonnqvist, T. Paunio, A. D. Børglum, A. Hartmann, A. Fink-Jensen, M. Nordentoft, D. Hougaard, B. Norgaard-Pedersen, Y. Böttcher, J. Olesen, R. Breuer, H. J. Möller, I. Giegling, H. B. Rasmussen, S. Timm, M. Mattheisen, I. Bitter, J. M. Réthelyi, B. B. Magnusdottir, T. Sigmundsson, P. Olason, G. Masson, J. R. Gulcher, M. Haraldsson, R. Fossdal, T. E. Thorgeirsson, U. Thorsteinsdottir, M. Ruggeri, S. Tosato, B. Franke, E. Strengman, L. A. Kiemeney; Genetic Risk and Outcome in Psychosis (GROUP); I. Melle, S. Djurovic, L. Abramova, V. Kaleda, J. Sanjuan, R. de Frutos, E. Bramon, E. Vassos, G. Fraser, U. Ettinger, M. Picchioni, N. Walker, T. Toulopoulou, A. C. Need, D. Ge, J. L. Yoon, K. V. Shianna, N. B. Freimer, R. M. Cantor, R. Murray, A. Kong, V. Golimbet, A. Carracedo, C. Arango, J. Costas, E. G. Jönsson, L. Terenius, I. Agartz, H. Petursson, M. M. Nöthen, M. Rietschel, P. M. Matthews, P. Muglia, L. Peltonen, D. St Clair, D. B. Goldstein, K. Stefansson, and D. A. Collier (2009): “Common variants conferring risk of schizophrenia,” Nature, 460, 744–747.

Tibshirani, R. (1996): “Regression shrinkage and selection via the lasso,” J. R. Stat. Soc. Series B (Methodological), 58, 267–288.

Tibshirani, R. (2011): “Regression shrinkage and selection via the lasso: a retrospective,” J. R. Stat. Soc. Series B (Statistical Methodology), 73, 273–282.

Tibshirani, R. J. (2013): “The lasso problem and uniqueness,” Electron. J. Stat., 7, 1456–1490.

Valdar, W., J. Sabourin, A. Nobel and C. C. Holmes (2012): “Reprioritizing genetic associations in hit regions using LASSO-based resample model averaging,” Genet. Epidemiol., 36, 451–462.

Waldron, L., M. Pintilie, M.-S. Tsao, F. A. Shepherd, C. Huttenhower and I. Jurisica (2011): “Optimized application of penalized regression methods to diverse genomic data,” Bioinformatics, 27, 3399–3406.

Warren, H., J.-P. Casas, A. Hingorani, F. Dudbridge and J. Whittaker (2014): “Genetic prediction of quantitative lipid traits: comparing shrinkage models to gene scores,” Genet. Epidemiol., 38, 72–83.

Willemsen, G., J. M. Vink, A. Abdellaoui, A. den Braber, J. H. van Beek, H. H. Draisma, J. van Dongen, D. van ‘t Ent, L. M. Geels, R. van Lien, L. Ligthart, M. Kattenberg, H. Mbarek, M. H. de Moor, M. Neijts, R. Pool, N. Stroo, C. Kluft, H. E. Suchiman, P. E. Slagboom, E. J. de Geus and D. I. Boomsma (2013): “The adult netherlands twin register: twenty-five years of survey and biological data collection,” Twin Res. Hum. Genet., 16, 271–281.

Zhang, P. (1993): “Model selection via multifold cross validation,” Ann. Stat., 21, 299–313.