

# Comparison of Lasso and Stepwise Regression in Psychological Data

Di Jody Zhou<sup>1</sup> , Rajpreet Chahal<sup>2</sup> , Ian H. Gotlib<sup>2</sup> , Siwei Liu<sup>1</sup> 

[1] Department of Human Ecology, University of California, Davis, CA, USA. [2] Department of Psychology, Stanford University, Stanford, CA, USA.

---

Methodology, 2024, Vol. 20(2), 121–143, <https://doi.org/10.5964/meth.11523>

Received: 2023-03-08 • Accepted: 2024-05-31 • Published (VoR): 2024-06-28

Handling Editor: Shahab Jolani, Maastricht University, Maastricht, the Netherlands

Corresponding Author: Di Jody Zhou, Department of Human Ecology, University of California, Davis, CA, USA. E-mail: [jodzhou@ucdavis.edu](mailto:jodzhou@ucdavis.edu)

Supplementary Materials: Code, Materials



## Abstract

Identifying significant predictors of behavioral outcomes is of great interest in many psychological studies. Lasso regression, as an alternative to stepwise regression for variable selection, has started gaining traction among psychologists. Yet, further investigation is valuable to fully understand its performance across various psychological data conditions. Using a Monte Carlo simulation and an empirical demonstration, we compared Lasso regression to stepwise regression in typical psychological datasets varying in sample size, predictor size, sparsity, and signal-to-noise ratio. We found that: (1) Lasso regression was more accurate in within-sample selection and yielded more consistent out-of-sample prediction accuracy than stepwise regression; (2) Lasso with a harsher shrinkage parameter was more accurate, parsimonious, and robust to sampling variability than the prediction-optimizing Lasso. Finally, we concluded with cautious notes and recommendations in practice on the application of Lasso regression.

## Keywords

variable selection, Lasso regression, stepwise regression, Monte Carlo simulation, model comparison

For many years, social and behavioral scientists have been advised against using stepwise regression methods due to several shortcomings concerning their variable selection accuracy and predictive ability (Thompson, 1995). Specifically, stepwise regression based on null hypothesis testing is prone to inflated false positive rates due to multiple testing (Derksen & Keselman, 1992), and the selected model might not include the most infor-



mative predictors as it is heavily influenced by the selection order in which predictors are added or removed (Lovell, 1983; Wilkinson, 1979). Moreover, the selection criteria used in stepwise regression are often too liberal, producing overly complex models with unnecessary predictors that perform poorly on new, unseen data, a phenomenon known as overfitting (Babyak, 2004). Overfitting is commonly seen in psychology research when the primary focus is to describe behavioral patterns within a specific theoretical framework (Yarkoni & Westfall, 2017). Most studies follow the conventional wisdom of quantitative analysis to select models that best approximate the characteristics of the data at hand (Davis-Stober et al., 2018). However, the best-fitted model does not guarantee reliable and replicable findings (Yarkoni & Westfall, 2017). This is because well-fitted models that closely capture all information in the data, often includes random fluctuations specific to a particular sample. The resulting failure of out-of-sample prediction contributes to the current replicability crisis in the behavioral sciences, in which replication studies often produce null results or smaller effects than the original studies (Laws, 2016).

In addressing this problem, some researchers have advocated for adopting machine learning approaches in psychology, which optimize prediction in new samples (Yarkoni & Westfall, 2017). According to this perspective, to enhance research replicability, it is pivotal to shift the analytical attention from mainly on explanation to a balance between explanation and prediction (Shmueli, 2010). In the context of variable selection, the goal is thus to select variables that are predictable in different samples of the same population and to disregard those that only capture nuances of a specific sample. Two approaches aiming at reducing model complexity can reduce overfitting and optimize model predictability. One is to reduce the number of predictors selected by stepwise regression using information-based fit indices, such as the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978). The other applies regularization on coefficient estimates through, for example, the least absolute shrinkage and selection operator (Lasso; Tibshirani, 1996). This study aims to investigate Lasso's performance as a variable selection method compared to stepwise regression in typical psychological data. In the following sections, we review the details of each approach.

## Stepwise Regression Using the F-Test

Traditionally, variable selection in stepwise regression is carried out based on null hypothesis testing, such as the F-test. In forward selection (i.e., selection starts with an empty model or just the intercept), for example, the  $p$ -value for each candidate predictor at each step  $t$  is calculated based on the partial  $F$  statistic as follows:

$$\Delta F = \frac{RSS_t - RSS_{t-1}}{RSS_{t-1}} \sum_{i=1}^N \left( y_i - \sum_{j=1}^p \beta_j \times x_{ij} \right)^2 \quad (1).$$

Each step selects the predictor with the smallest  $p$ -value. This iterative algorithm terminates when no remaining predictors meet the inclusion threshold (i.e., a prespecified  $\alpha$  value). The magnitude of this  $\alpha$  value determines the selection liberty: a higher  $\alpha$  allows for the inclusion of more predictors but increases the false positive rate. Backward elimination does the opposite of forward selection. Starting with a full model that includes all variables, each step prunes a variable until no remaining predictor has a  $p$ -value larger than the prespecified inclusion threshold. Both-directional stepwise regression combines forward selection and backward elimination by either adding a variable or pruning a variable at each step; a variable that enters the model at a prior step can also be removed later. Regardless of the direction of variable selection, these methods aim to arrive at a final model that best fits the given data.

## Stepwise Regression With Information-Based Fit Indices

Information-based fit indices penalize the number of predictors included in the stepwise regression to maximize model fit with the fewest predictors<sup>1</sup>. The two most common indices, namely, AIC and BIC, differ in the degree of penalty they place on the number of predictors. AIC has a fixed penalty of 2, whereas the penalty level of BIC,  $\ln(N)$ , increases with the sample size. As  $\ln(N)$  is typically larger than 2 in psychological research, BIC generally favors more parsimonious models than AIC. Specifically, at each step  $t$ , the difference in AIC and BIC is calculated as follows:

$$\Delta AIC = N \times \ln\left(\frac{RSS_t}{N}\right) - N \times \ln\left(\frac{RSS_{t-1}}{N}\right) + 2 \times (K_t - K_{t-1}) \quad (2),$$

$$\Delta BIC = N \times \ln\left(\frac{RSS_t}{N}\right) - N \times \ln\left(\frac{RSS_{t-1}}{N}\right) + \ln(N) \times (K_t - K_{t-1}) \quad (3),$$

where  $K$  is the number of predictors to be estimated in the model. Hence,  $K_t - K_{t-1}$  is 1 for forward selection and -1 for backward elimination.

## Lasso Regression

The Lasso regression aims to reduce overfitting with regularization, a general approach to reduce prediction error in unseen samples by finding a balance between bias and variance. In the context of regression, bias refers to the distance of the estimated coefficients from the true coefficients, and variance refers to the uncertainty of the coefficient estimates due to sampling variability. The two elements often compete with

---

1) Despite the addition of the penalty terms, prior studies have shown that stepwise regression methods with information-based fit indices perform similarly to traditional stepwise regression methods based on hypothesis testing (Heinze et al., 2018; Sauerbrei, 1999).

each other, a phenomenon known as the bias-variance tradeoff (Helwig, 2017; Liu & Rhemtulla, 2022; McNeish, 2015). Out-of-sample prediction errors are affected by both bias and variance. Hence, if the analytical goal is to optimize predictive performance, arriving at a model with minimal bias may not be ideal; instead, it may be helpful to enlarge bias for smaller variance. In regression, this is done by shrinking the coefficient estimates towards zero. Popular regularization methods include Ridge regression, Lasso regression, and Elastic-net methods. Among them, Lasso regression is arguably the most frequently used variable selection method (Tibshirani, 1996; Zou et al., 2007). In addition to minimizing the *RSS*, Lasso adds a penalty on the sum of the absolute values of the beta coefficients:

$$\hat{\beta}^{\text{lasso}} = \operatorname{argmin} \left\{ \text{RSS} + \lambda \times \sum_{j=1}^p |\beta_j| \right\} \quad (4).$$

This constraint attenuates the coefficients of all predictors, with the degree of attenuation determined by the tuning parameter  $\lambda$ . As  $\lambda$  increases, the coefficients of predictors with small effects on the outcome variable will be quickly shrunk to zero, whereas the coefficients of predictors with larger effects will remain non-zero.

As different  $\lambda$  selects different sets of variables, an important task in Lasso regression is to select the optimal value of  $\lambda$  to reach a desired level of model sparsity. A too-small  $\lambda$  would include almost all variables. A too-large  $\lambda$  can result in an under-fitted model with too few predictors. The information-based fit indices AIC or BIC (Zou et al., 2007) and the k-fold cross-validation method (Golub et al., 1979) can be used to determine an appropriate level of  $\lambda$ . Among these methods, k-fold cross-validation is used more frequently because information-based fit indices require the calculation of degrees of freedom, which is challenging (McNeish, 2015; Tibshirani & Taylor, 2012). K-fold cross-validation splits the dataset into k samples. Each sample takes turns validating the model trained by the rest of the k-1 samples. For each iteration of training and validation, a series of the  $\lambda$  values are used to generate and compare the cross-validation mean-squared error (*MSE*)<sup>2</sup>. The  $\lambda$  value that minimizes the cross-validation *MSE* (hereinafter denoted as  $\lambda_{\min}$ ) is recommended for variable selection. Alternatively, others recommend using the largest  $\lambda$  value that produces a cross-validation *MSE* within one standard error of the minimum cross-validation *MSE* (hereinafter denoted as  $\lambda_{1se}$ ) for better model parsimony (Friedman et al., 2010; McNeish, 2015).

Lasso regression has been recently introduced to psychologists (Helwig, 2017; Jacobucci et al., 2019; McNeish, 2015), and has been gaining traction in empirical analyses to enhance the prediction of psychological behaviors (Haynos et al., 2021; Smith

2) The relation between  $\lambda$  and the average *MSE* across iterations typically follows a U-shape pattern because underfitted and overfitted models would both produce large cross-validation errors.

et al., 2020). Yet, most methodological studies assessing its performance focused on high-dimensional data (i.e., the number of predictors is far larger than the sample size; Bickel et al., 2009; Fan & Lv, 2010; Sirimongkolkasem & Drikvandi, 2019). In contrast, psychological datasets often contain more observations than variables. Although a few recent studies have found that Lasso is more accurate in predicting the hold-out sample than stepwise regression when applied to low-dimensional datasets (Ahrens et al., 2020; Greenwood et al., 2020; Xu et al., 2012), others have shown that this improvement in prediction is very small and limited to a few conditions (Hastie et al., 2020; Pavlou et al., 2016; Wester et al., 2022). The differences in data characteristics are the primary reason for these mixed findings. Among these studies, two investigated categorical response variables from a single empirical dataset (Greenwood et al., 2020; Xu et al., 2012), whereas others included continuous response variables in simulated datasets (Ahrens et al., 2020; Hastie et al., 2020; Wester et al., 2022). Moreover, their simulated data did not well represent characteristics of data commonly seen in psychological research. For example, Ahrens et al. (2020) and Hastie et al. (2020) examined datasets with 100 candidate predictors, which is rarely the case in psychological research. In Wester et al. (2022), which focuses on the selection of interactions to model treatment effect heterogeneity, relevant and irrelevant predictors were simulated to be independent. However, this assumption is unrealistic, given that correlated covariates are ubiquitous in psychological studies. Therefore, it is still unclear how this novel Lasso regression method compares to the more popular stepwise regression method when applied to psychological data. Given the increasing popularity of the Lasso method in psychology, more research is needed to better understand its performance in typical psychological data, especially its selection accuracy, which Lasso regression was not developed to optimize but is highly desired in psychological research to draw inferential conclusions.

To address this issue, we conducted a Monte Carlo simulation study to compare Lasso and stepwise regression across a representative range of data conditions that are typical in psychological studies. We organize the remainder of this article as follows. First, we describe our simulation study comparing Lasso and stepwise regression in out-of-sample prediction, within-sample selection accuracy, and model sparsity. Next, we demonstrate their differences in an empirical study that aims to identify risk factors of adolescent externalizing problems. Finally, we conclude with cautious notes and recommendations about the application of these methods in practice.

## Method

### Simulation Design

Based on our literature review of recent psychological studies that involve variable selection methods (see Supplementary Material A in Zhou et al., 2024), we simulated data

using a four-way factorial design. Specifically, we manipulated: 1) sample size ( $N$ ), which can be 100, 200, 400, or 800; 2) candidate predictor size ( $p$ ), which can be 5, 15, 25, 35, or 80<sup>3</sup>; 3) signal-to-noise ratio (SNR), defined as  $\frac{\text{Var}(\hat{y})}{\text{Var}(y - \hat{y})}$  or  $\frac{R^2}{1 - R^2}$ , can be 0.2, 0.5, 0.8, or 2; and 4) level of sparsity ( $s$ ), defined as the percentage of non-informative predictors, which can be 20%, 40%, or 80%. In total, the factorial design yielded 240 ( $4 \times 5 \times 4 \times 3$ ) conditions, with 100 replications of each simulation condition. All data were simulated based on the multiple linear regression model without the intercept:  $y = X\beta + \epsilon$ , where  $X$  is an  $N \times p$  matrix of candidate predictor values drawn from  $N_p(0, \Sigma)$ .  $\Sigma$  is a  $p \times p$  correlation matrix with the elements on the diagonal fixed to 1, and the other entries randomly and independently drawn from  $Beta(3.5, 3.5)$  bounded between -1 and 1. We used this distribution so that most correlations among the candidate predictors ranged between -0.5 and 0.5, but some candidate predictors were allowed to be highly correlated with each other. We ensured that the predictor correlation matrix was positive semi-definite by replacing a non-positive definite correlation matrix with the nearest positive-definite correlation matrix (Higham, 2002). In each replication, a unique correlation matrix  $\Sigma$  was used to generate the candidate predictor values.  $\beta = (\beta_1, \dots, \beta_p)^T$  contains regression coefficients with set  $J_1 = \{j, \beta_j = 1\}$  representing the informative predictors and set  $J_2 = \{j, \beta_j = 0\}$  the non-informative predictors.  $y = (y_1, \dots, y_n)^T$  is the outcome variable from  $N_n(X\beta, \sigma^2 I_n)$  where  $\sigma^2 = \frac{\text{Var}(X\beta)}{\text{SNR}}$ , and  $I_n$  is the  $n \times n$  identity matrix. Within each replication, we simulated a training set and a test set using the same parameter values. The test set was used to calculate the out-of-sample predictive accuracy, as described below.

## Simulation Analysis

The *stepAIC* function in the *MASS* package was used (Venables & Ripley, 2002) to perform forward selection, backward elimination, and both-direction stepwise regression with three different selection criteria including the traditional  $F$ -statistic method with  $\alpha = 0.15$  (i.e., the threshold criteria compared to the  $p$ -value of each predictor) as recommended in the literature (Derksen & Keselman, 1992; Flack & Chang, 1987), and the information-based fit indices AIC and BIC. We then used the *cvglmnet* function in the *glmnet* package (Friedman et al., 2010) to conduct Lasso regression. Five-fold cross-validation was used to optimize the tuning parameter  $\lambda$ . We performed Lasso with both  $\lambda_{\min}$  and  $\lambda_{1se}$  (respectively denoted as Lasso.min and Lasso.1se below). Because Lasso is sensitive to the scale of the coefficient values during variable selection, all variables were standardized prior to the analysis (Hastie et al., 2009, Chapter 3)<sup>4</sup>.

3) Although  $p = 80$  is outside the range of data characteristics found in our literature review (see Supplementary Material A in Zhou et al., 2024), we included this condition to represent characteristics of “big data,” which is gaining momentum in studies with access to online data and team science work.

4) Previous research showed that Lasso does not select the true model asymptotically (i.e., when the sample size is sufficiently large) unless the *irrepresentable condition* (IRC) holds, in which the correlations between important and

## Comparison Criteria

### Out-of-Sample Predictive Accuracy

Predictive accuracy is measured by the out-of-sample *MSE*, which is the mean squared distance between the outcome values estimated by the train set and the true outcome values in the corresponding test set:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (5),$$

where  $\hat{y}_i$  is the *i*th outcome value estimated by the train set, and  $y_i$  is the *i*th true outcome value in the test set.

### Within-Sample Selection Accuracy

We examined three measures of selection accuracy in the training set, including sensitivity, specificity, and the Matthew correlation coefficient (MCC) (Baldi et al., 2000). They are calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \text{ Specificity} = \frac{TN}{TN + FP},$$

$$\text{and MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (6),$$

where true positive (*TP*) is the number of selected predictors that are truly informative; true negative (*TN*) is the number of eliminated predictors that are truly non-informative; false positive (*FP*) is the number of selected predictors that are truly informative in the true model, and false negative (*FN*) is the number of selected predictors that are not important in the true model.

Both sensitivity and specificity range from 0 to 1, with higher values indicative of higher selection rates of truly informative predictors and higher exclusion rates of non-informative predictors, respectively (Altman & Bland, 1994). MCC measures the overall selection accuracy of the estimated model by accounting for all four categories of the confusion matrix (i.e., *TP*, *TN*, *FP*, *FN*) in a single metric. We chose MCC as the

---

unimportant predictors are weak (Zhao & Yu, 2006). We investigated the effect of IRC and the interaction effect of IRC and selection method on selection accuracy and predictive performance. We found that IRC was associated with more accurate selection and prediction, but the interaction effect of IRC and selection method is very small ( $\eta^2 < 0.01$ ). The ANOVA results are in Table 4 of the Supplementary Material C of Zhou et al. (2024). This means that Lasso is not more or less susceptible to IRC than stepwise regression methods in the data conditions generated by our simulation. Therefore, we did not separate the IRC and non-IRC conditions when presenting our results as we would expect that both Lasso regression and stepwise regression perform better in the IRC conditions, but no method benefits more than others.

third comparing criterion for its comprehensive evaluation of the selection quality that captures the balance between sensitivity and specificity. This metric ranges between -1 and 1, with values close to 1 representing a strong positive correlation between the estimated model and the true model, and values close to -1 representing a strong negative correlation between the estimated model and the true model.

### Model Sparsity

Model sparsity measures model parsimony. It represents the percentage of non-selected predictors among all candidate predictors. This measure ranges from 0 to 1.

### Analysis

We conducted a factorial repeated-measures analysis of variance (RM-ANOVA; Myers, 1979) to investigate the main effects of predictor size, sample size, SNR, sparsity, selection method, and their interaction effects on each comparison criterion. We evaluated those effect sizes through the generalized eta squared ( $\eta_G^2$ ) (Olejnik & Algina, 2003). Because there were many terms in the ANOVA models, we only reported results with at least a small effect size (i.e.,  $\eta_G^2 \geq 0.01$ ; Cohen, 1988), and we focused on interaction effects that involve selection methods. All steps of simulation and analyses were conducted in R Version 4.0.0 (R Core Team, 2020), and the simulation code is in the Supplementary Material B of Zhou et al. (2024).

## Simulation Study Results

### Preliminary Analyses

All three backward elimination methods (i.e., backward elimination with  $F$ -statistic, AIC, and BIC) produced far more complex models with significantly poorer predictive performance and selection accuracy. For example, the  $MSE$  of all backward elimination methods with small sample sizes ( $N < 200$ ) were disproportionately high with values larger than 2000; the MCC of all backward elimination methods were almost 0 when predictor size was larger than 15; increase in sample size did not improve sensitivity as other methods did<sup>5</sup>. We also found that the forward selection methods performed very

---

5) A careful look at the data suggests that this was due to multicollinearity. For example, in one dataset with five predictors, a sample size of 100, sparsity as 0.4, and SNR as 0.2, the minimum variance inflated factor ( $VIF$ ) was 849. High  $VIF$  suggests a severe degree of multicollinearity. Moreover, pruning collinear variables of similar significance backward starting from a full model of all candidate predictors might exacerbate the selection of unnecessary variables, distorting the magnitudes of regression coefficients. The cross-validation  $MSE$  is thus extremely large. In some other programs (e.g., SPSS), candidate variables with  $VIF > 10$  are automatically excluded from the selection procedure (IBM Corp, 2020). The  $stepAIC$  function in R does not have this built-in feature. In this article, we decided to exclude backward elimination methods in the following comparison analyses because: 1) the sensitivity



similarly to the both-directional stepwise regression methods. Hence, we only report the results of both-directional stepwise regression methods below<sup>6</sup>. The descriptive statistics for all methods are included in Supplementary Material C of Zhou et al. (2024), and Table 5 in Supplementary Material C of Zhou et al. (2024) details the ANOVA results for the both-directional stepwise regression methods and the Lasso methods.

### Out-of-Sample Predictive Accuracy

There was no sizeable difference in *MSE* across conditions or methods, as none of the main or interactive effects had an effect size  $\eta^2$  larger than 0.01. The descriptive statistics (see Table 1) show that Lasso.min returns with the lowest prediction error, although its difference with stepwise regression, especially with Stepwise.bic, is very small. The standard deviations of the two Lasso methods were smaller than most stepwise regression methods, except for being almost identical as Stepwise.bic, indicating a more consistent predictive accuracy than most stepwise regression methods. Stepwise.bic was the most accurate and consistent stepwise regression method.

**Table 1**

*MSE Across Methods*

<i>MSE</i>	Lasso.1se	Lasso.min	Stepwise.aic	Stepwise.bic	Stepwise.f
<i>M</i>	0.68	0.64	6.76	0.66	3.15
<i>Mdn</i>	0.69	0.65	0.66	0.67	0.66
<i>SD</i>	0.21	0.20	317	0.21	130

### Within-Sample Selection Accuracy

**Sensitivity** — Figure 1 depicts the levels of sensitivity across conditions and methods. Only sample size of 100 is shown here for demonstration because, 1) it is the closest to

---

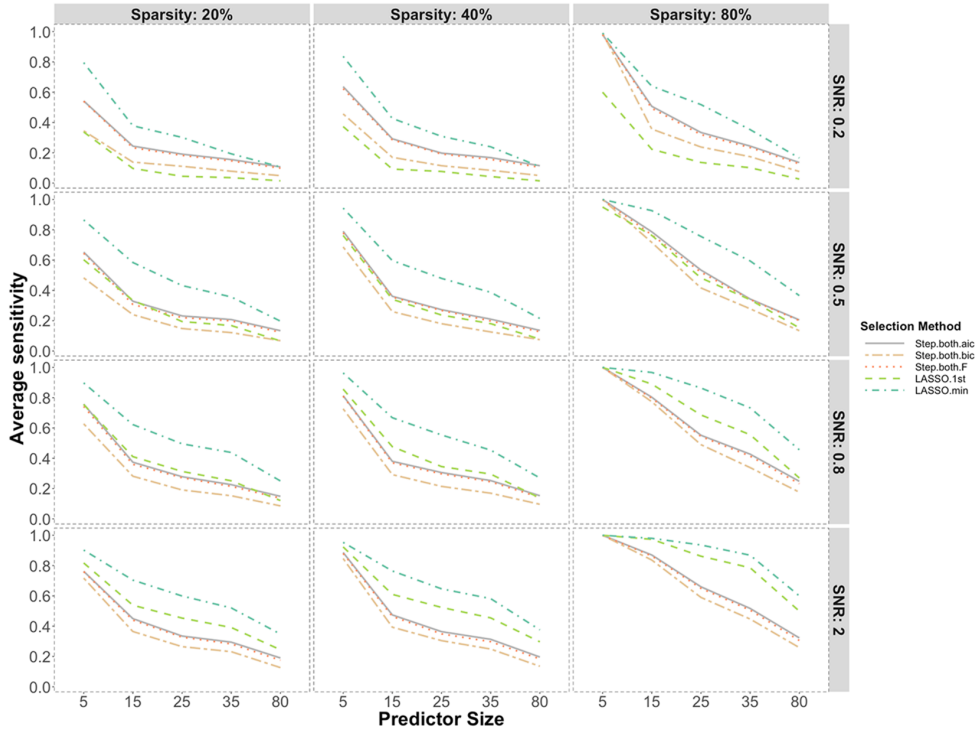
to multicollinearity is a unique problem related to the *stepAIC* function we used, not a universal problem across statistical programs; 2) the problem can be solved by applying prescreening methods such as those in SPSS. Interested readers can find the results of the backward elimination methods in the Supplementary Material C of Zhou et al. (2024). We also address the issue of multicollinearity in the discussion.

6) The recent literature has recommended the integration of cross-validation into stepwise regression for better performance (Hastie et al., 2020 and Wester et al., 2022). We conducted forward selection and backward elimination with 10-fold cross-validation following Wester et al. (2022) to optimize predictor size that is associated with the best out-of-sample performance. While forward selection with cross-validation still does not outperform Lasso in within-sample selection and out-of-sample prediction, backward elimination benefits substantially from cross-validation for a better selection and prediction accuracy. However, it still performs worse than forward and both-directional stepwise regression. We did not include them in our main analysis because stepwise regression with cross-validation has not been widely used by psychologists due to the lack of available statistical software options.

our empirical data sample size, and 2) sample size has the least influence on all measures without imposing any interaction effects with the analytical method.

**Figure 1**

*Average Sensitivity Across Methods and Across Conditions When Sample Size Was 100*



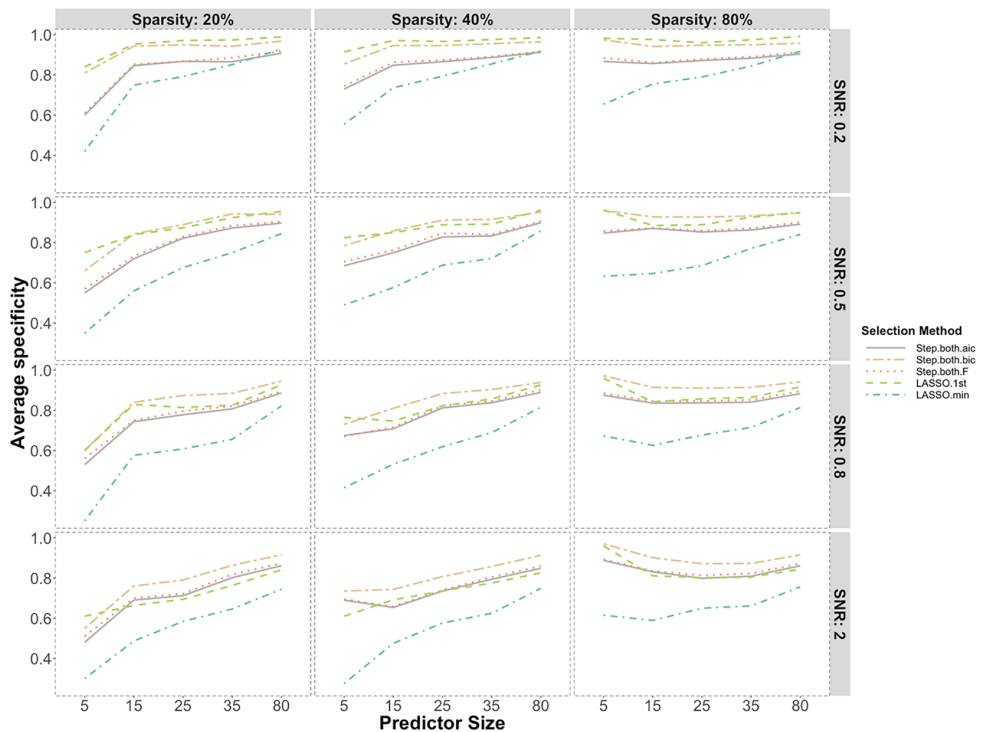
The results of different sample sizes are included in the Supplementary Material C of Zhou et al. (2024) for interested readers. There were large main effects of predictor size ( $\eta^2 = 0.65$ ), sparsity ( $\eta^2 = 0.4$ ), selection method ( $\eta^2 = 0.26$ ), SNR ( $\eta^2 = 0.19$ ), and sample size ( $\eta^2 = 0.15$ ). There were also small interaction effects between SNR and selection method ( $\eta^2 = 0.03$ ), between predictor size and selection method ( $\eta^2 = 0.03$ ), and a three-way interaction effect between predictor size, sparsity, and selection method ( $\eta^2 = 0.02$ ). Specifically, sensitivity increased as sparsity, SNR, and sample size increased (see Table 2.2 in the Supplementary Material C of Zhou et al., 2024) but decreased as predictor size increased. The Lasso methods, especially Lasso.min had the greatest sensitivity across all conditions. Lasso.1st had higher sensitivity than stepwise regression unless SNR was as small as 0.2. This means that the Lasso methods were generally better at identifying the informative predictors than the stepwise regression methods unless the data were very

noisy. Furthermore, the difference in sensitivity between Lasso and stepwise regression was minimal when only a few candidate predictors were considered ( $p = 5$ ), and only a few were informative ( $s = 0.8$ ). This difference was larger when predictor size and the number of informative predictors increased.

**Specificity** — Similar to sensitivity, all of the five factors of interest: selection method ( $\eta^2 = 0.17$ ), predictor size ( $\eta^2 = 0.07$ ), sparsity ( $\eta^2 = 0.06$ ), SNR ( $\eta^2 = 0.04$ ), and sample size ( $\eta^2 = 0.02$ ), had at least small effects on specificity. The interaction between predictor size and selection method yielded a small effect ( $\eta^2 = 0.02$ ). Figure 2 depicts the levels of specificity across conditions and methods.

**Figure 2**

*Average Specificity Across Methods and Across Conditions When Sample Size Was 100*



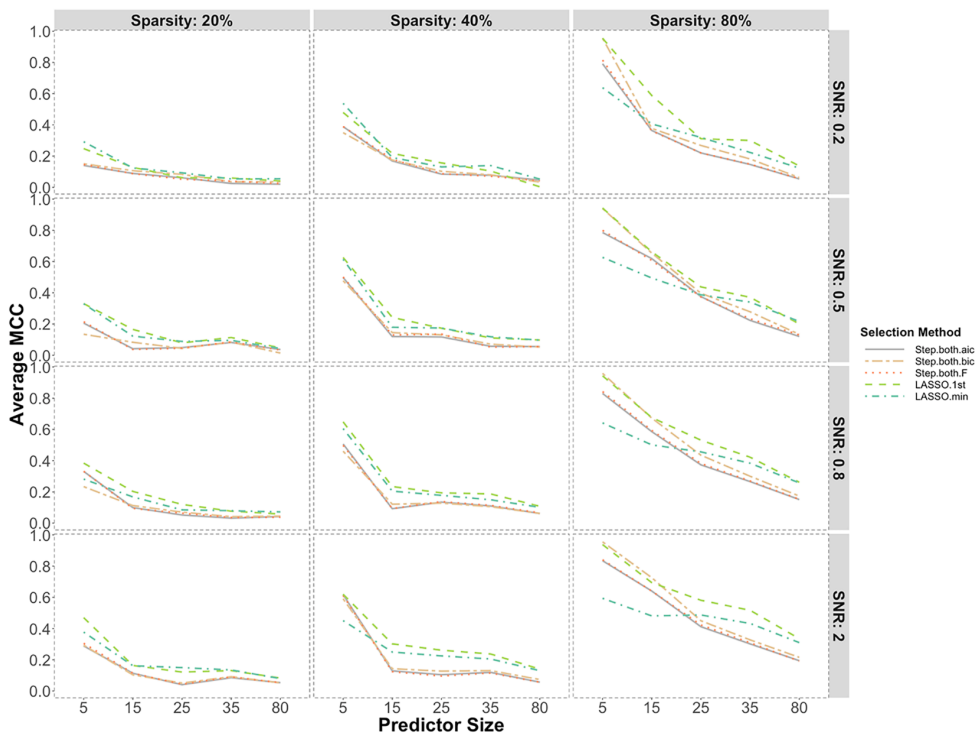
Larger predictor sizes and sparsity levels, lower SNR, and smaller sample sizes (see Table 2.3 in the Supplementary Material C of Zhou et al., 2024) were associated with a higher level of specificity. Lasso.min had the lowest specificity across all conditions, and Lasso.1se outperformed stepwise regression methods only when SNR was lower than 0.5.

This indicates that Lasso.min was more likely to classify non-informative predictors as significant and thus produced models with larger false positive or Type 1 error rate than all others. Lasso.1se was better but did not outmatch stepwise regression methods unless in conditions wherein only a small proportion of the outcome variance was explainable by the linear regression. Moreover, the difference in specificity across methods was magnified in conditions with only five predictors. For example, Lasso.min’s degree of specificity was almost twice as small as most other methods, and the specificity of Lasso.1se was the highest.

**MCC** – Sparsity level ( $\eta^2 = 0.3$ ) and predictor size ( $\eta^2 = 0.22$ ) had a large effect size when predicting MCC. Selection method ( $\eta^2 = 0.02$ ) and the three-way interaction effect between predictor size, sparsity, and selection method ( $\eta^2 = 0.02$ ) exhibited small effects on MCC. **Figure 3** depicts the levels of MCC across conditions and methods.

**Figure 3**

*Average MCC Across Methods and Across Conditions When Sample Size Was 100*

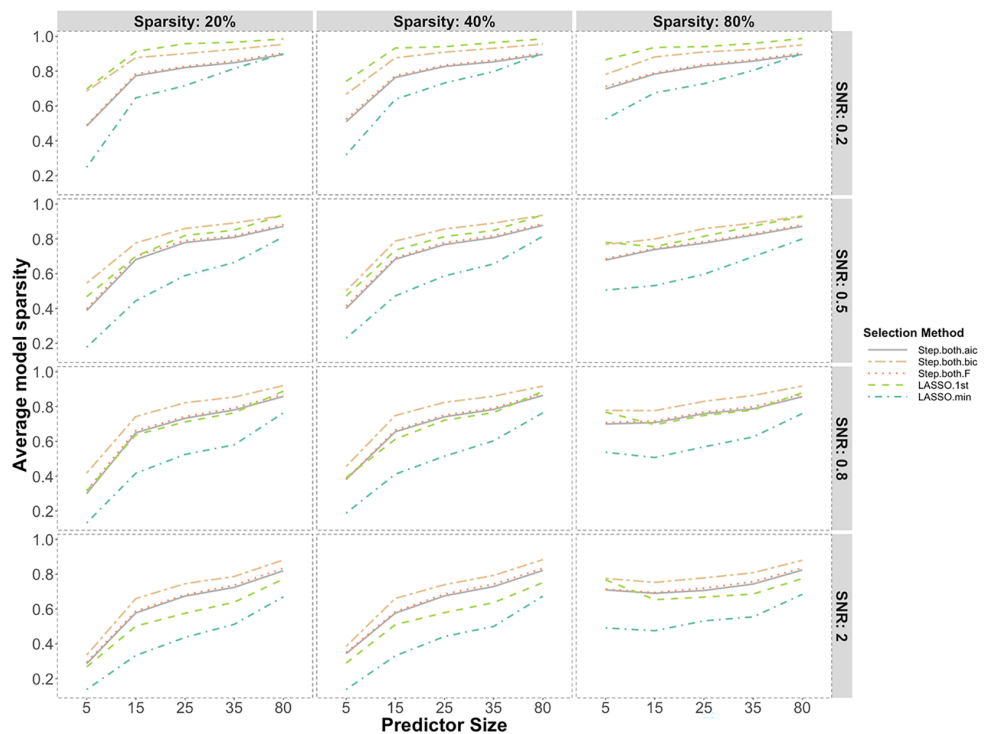


In general, MCC increased with higher sparsity levels and smaller candidate predictor sizes. This implies that selection is generally more accurate when dealing with a small pool of candidate variables of which only a few are truly informative. Across methods, although Lasso.min had the lowest MCC when predictor size was small ( $p < 25$ ) and only a few were informative ( $s = 0.8$ ), Lasso regression methods had higher MCC across most other conditions. In particular, Lasso.1se had the highest MCC across all conditions. This indicates that Lasso regression methods, especially Lasso with a harsher shrinkage tuning parameter, were more likely to produce models with better selection accuracy than stepwise regression methods.

**Model Sparsity** – All the five main effects yielded large effect size: predictor size ( $\eta^2 = 0.63$ ), selection method ( $\eta^2 = 0.47$ ), sparsity ( $\eta^2 = 0.26$ ), SNR ( $\eta^2 = 0.24$ ), and sample size ( $\eta^2 = 0.17$ ). There were also two small-size interaction effects with the selection method: SNR and selection method ( $\eta^2 = 0.04$ ), and predictor size and selection method ( $\eta^2 = 0.03$ ). **Figure 4** depicts differences in model sparsity across conditions and methods.

**Figure 4**

*Average Model Sparsity Across Methods and Across Conditions When Sample Size Was 100*



Model sparsity generally increased when predictor size and sparsity level increased; it decreased when sample size (see Table 2.5 in Supplementary Material C of Zhou et al., 2024) and SNR increased. This means that for all methods, a more parsimonious model was more likely to be obtained if the candidate predictor pool was large, a few predictors were truly relevant, the sample size was small, and the data were noisy. Between methods, Lasso.min produced the least parsimonious models across all conditions. On the other hand, stepwise regression with BIC produced more parsimonious models than all other methods in most conditions except when the SNR was 0.2 or when the predictor size was 5. In such cases, Lasso.1se produced models similar in size or more parsimonious than stepwise regression with BIC.

## Empirical Example

### The Empirical Data

We demonstrate the variable selection differences between Lasso and stepwise regression in an empirical study aimed at predicting externalizing behavior (i.e., disruptive, aggressive, or delinquent actions that are directed outwardly) during adolescence. The data were drawn from the second timepoint of the Stanford Early Life Stress Study (Chahal et al., 2022; Gotlib et al., 2021), where externalizing behavior is assessed by the aggressive behavior and rule-breaking behavior scales of the Youth Self-Report (YSR; Achenbach, 2001), a widely used measure of behavioral problems in adolescents. Thirty-three externalizing-behavior-relevant variables from the domains of pubertal development, sensitivity to stress and reward, emotional and behavioral problems, social support, physical and emotional neglect, emotional and affective regulation, early life stress severity, and demographic information of the child and parents, were pre-selected as predictors of externalizing behavior. Complete case analysis was conducted with a sample size of 141. A detailed description of the measures of the predictors and the sample is included in the Supplementary Material D of Zhou et al. (2024).

### Analysis

To examine the predictive performance of the models, we split the full dataset into a training set with 80% of the sample and a test set with the remaining 20%. This random splitting procedure was repeated 1000 times to investigate selection variability due to sample variation. Methods were compared in terms of their average out-of-sample predictive accuracy (i.e., *MSE*), average model sparsity (i.e., the percentage of variables not selected across the 1000 iterations), and selection rate of each variable.

## Empirical Study Results

### Model Predictive Accuracy and Model Sparsity Across Methods

Table 2 shows the statistical summary of out-of-sample *MSE*, model size (i.e., number of predictors selected), estimated model sparsity, and SNR for each method across the 1000 iterations of model training and testing. Similar to the simulation results, methods did not differ much in the out-of-sample predictive accuracy. Stepwise regression with BIC yielded the most parsimonious model, and it was more consistent in model size than other methods, as indicated by a smaller standard deviation. On the other hand, Lasso.min tended to produce the most complicated model and varied most in model size, as indicated by a larger standard deviation. Because the true model was unknown, it was impossible to compare across methods in selection accuracy. However, based on our simulation results on the average MCC for a sample size of 100 (see Figure 3), we would expect the selection accuracy of Lasso.1se to be the best given the characteristics of the data (i.e., predictor size = 33; sample size = 113; estimated sparsity level ranging from 0.55 to 0.84; estimated SNR ranging from 0.64 to 2.14).

### Summary of Selection

Figure 5 shows the selection rate of each candidate predictor. Internalizing problems, impulsivity, affective reactivity, sex (being male), and sensitivity to punishment were selected most often by all methods. Among these methods, stepwise regression with BIC, stepwise regression with *F* statistic, and Lasso.1se tended to produce more parsimonious models than stepwise regression with AIC and Lasso.min. Moreover, they were more consistent in their selection, an indicator that these three methods were more robust to sample variability. It should be noted that, however, this discrepancy in selection consistency is confounded by model sparsity.

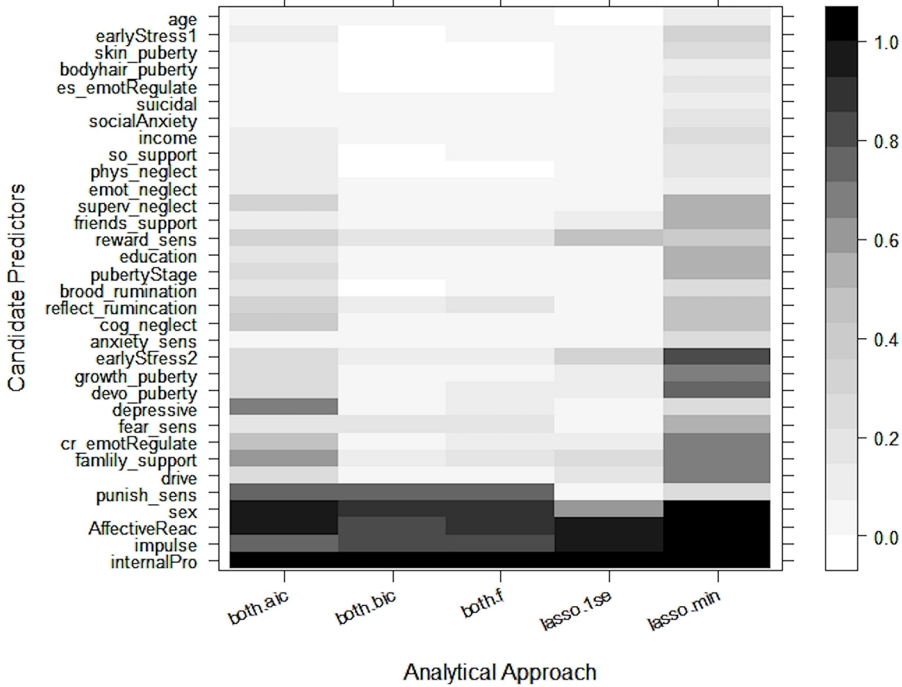
**Table 2**

*Summary of Method Performance Across 1000 Iterations*

Selection Method	MSE		Model Size		Estimated Sparsity		Estimated Snr	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Lasso.1se	0.53	0.11	5.54	2.96	0.83	0.09	0.64	0.28
Lasso.min	0.51	0.13	15	6.48	0.55	0.2	1.47	0.49
Stepwise.aic	0.56	0.15	10.4	2.27	0.69	0.07	2.14	0.33
Stepwise.bic	0.5	0.15	5.31	0.91	0.84	0.03	1.69	0.24
Stepwise.f	0.5	0.15	5.78	1.04	0.83	0.03	1.75	0.25

Figure 5

Selection Rate of Each Candidate Variable Across Methods



### Discussion

Our study evaluated the variable selection performance of Lasso regression in comparison to the most commonly used stepwise regression in psychological data. We found Lasso regression a competitive alternative to stepwise regression for their more accurate selection and more consistent out-of-sample prediction across different scenarios. However, Lasso’s improvement in minimizing prediction errors was negligible ( $\eta^2 < 0.001$ ), which is consistent with findings of previous simulation studies in low-dimensional data (Pavlou et al., 2016; Wester et al., 2022; Williams et al., 2019; Williams & Rodriguez, 2020). In terms of model sparsity, Lasso regression did not always yield more parsimonious models than stepwise regression unless the SNR was low (i.e., SNR is 0.2 or  $R^2$  is 0.17) and the candidate predictor pool was small (i.e., only five predictors). Stepwise.bic was overall the best stepwise regression, but others produced far more complex models and inconsistent prediction errors, which align well with the literature concerning their unreliable selection results. Finally, results from our empirical study are consistent with



the current psychopathology development literature, where internalizing problems are most salient for explaining externalizing problems (Chahal et al., 2022).

Our findings also corroborate previous research on the tradeoff between sensitivity and specificity (Su et al., 2017). Although Lasso regression is better at identifying the truly “significant” predictors of the outcome (i.e., higher sensitivity), its ability to exclude non-informative predictors is worse (i.e., lower specificity or higher false positive rate) than stepwise regression, especially when predictor size is very small and  $\lambda_{1se}$  is used for selection (Freijeiro-González et al., 2022; Greenwood et al., 2020; Guo et al., 2015; Pavlou et al., 2016; Xu et al., 2012). Between the two Lasso regression methods, the Lasso regression with a harsher penalty is known to produce models with a more balanced combination of power and false positive rate, leading to better overall selection accuracy (Hastie et al., 2020). We found this disparity most prominent when the predictor size was small and only a few were truly informative. In addition, we also found that the larger penalty parameter resulted in higher selection consistency (Meinshausen & Bühlmann, 2006) and a comparable model size as the stepwise regression with BIC (Morozova et al., 2015) produced the simplest models across most conditions.

## Cautious Notes and Recommendations in Practice

We recommend researchers to carefully consider their analytical goals and data characteristics before choosing a variable selection method. Lasso regression is generally more consistent in out-of-sample prediction, although its improvement in magnitude in typical psychological data is not as pronounced as it is in high-dimensional datasets. Lasso regression is also a competitive alternative to stepwise regression if the aim is to make accurate explanation for its better within-sample selection accuracy. However, neither of the two Lasso methods investigated here exhibited sufficiently high selection accuracy across all conditions. As shown in Figure 3, even for Lasso.1se, the one with the highest MCC, its MCC hardly exceeded 0.4 unless the predictor size was very small ( $p = 5$ ) or the sparsity level was very high ( $s = 0.8$ ). Because the sparsity level of the true model is unknown in empirical data, researchers may consider prescreening predictors based on theoretical knowledge or prior empirical evidence to narrow down the candidate pool to a size smaller than 15, if possible.

Given the tradeoff between power and sensitivity in variable selection, the ideal method should consider the analytical priority, such as to capture true effects or to avoid spurious findings. If statistical power is more of a concern, we recommend Lasso regression with the harsher penalty term ( $\lambda_{1se}$ ) for lower false positive rate, better overall selection accuracy, lower model complexity, and more consistent selection. On the other hand, when the priority is to reach lower false positive rate and to improve model interpretability, Lasso does not yield better results than stepwise regression with information-based fit indices, especially the BIC, particularly when the predictor size is small and only a small fraction of the predictors is truly relevant. To improve the false posi-

tive rate of the Lasso regression, we encourage researchers to consider more advanced derivatives such as the adaptive Lasso (Fan & Li, 2001; Zou, 2006) and the relaxed Lasso (Meinshausen, 2007). These methods further exclude less influential variables through their weighted degree of shrinkage that imposes a magnified penalty on non-informative predictors.

Following the machine learning literature, we recommend researchers to check the IRC assumption to evaluate the selection consistency when using Lasso (Meinshausen & Bühlmann, 2006; Zhao & Yu, 2006). It is important to note that the presence of multicollinearity—a common issue in psychological data where predictors are highly linearly dependent—can worsen the selection consistency of Lasso regression. This is because Lasso randomly selects one variable out of a group of highly correlated ones to reach a sparser model (Zou & Hastie, 2005). If researchers are interested in delivering a consistent model that contains all highly correlated predictors to include both interaction and main effects, derivatives of Lasso such as the elastic-net (Zou & Hastie, 2005) and the grouped Lasso (Yuan & Lin, 2006) are generally recommended. Another remedy for the unstable selection is to integrate Bayesian selection into variable selection. For example, the stochastic search variable selection (SSVS) assigns higher prior probability to more promising predictors (Bainter et al., 2020).

---

**Funding:** The authors have no funding to report.

---

**Acknowledgments:** We thank Dr. Donald R. Williams for his insightful comments and for drafting the initial R code to check the IRC assumption.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

## References

- Achenbach, T. M. (2001). *Manual for ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth & Families.
- Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020). lassopack: Model selection and prediction with regularized regression in Stata. *Stata Journal*, 20(1), 176–235. <https://doi.org/10.1177/1536867X20909697>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ (Clinical Research Ed.)*, 308, Article 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
- Babiyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421. <https://doi.org/10.1097/01.psy.0000127692.23278.a9>

- Baldi, P. A., Brunak, S., Chauvin, Y., Andersen, C. A. F., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics*, *16*(5), 412–424. <https://doi.org/10.1093/bioinformatics/16.5.412>
- Bainter, S. A., McCauley, T. G., Wager, T., & Losin, E. A. R. (2020). Improving practices for selecting a subset of important predictors in psychology: An application to predicting pain. *Advances in Methods and Practices in Psychological Science*, *3*(1), 66–80. <https://doi.org/10.1177/2515245919885617>
- Bickel, P. J., Ritov, Y. A., & Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, *37*(4), 1705–1732. <https://doi.org/10.1214/08-AOS620>
- Chahal, R., Miller, J. G., Yuan, J. P., Buthmann, J. L., & Gotlib, I. H. (2022). An exploration of dimensions of early adversity and the development of functional brain network connectivity during adolescence: Implications for trajectories of internalizing symptoms. *Development and Psychopathology*, *34*(2), 557–571. <https://doi.org/10.1017/S0954579421001814>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Davis-Stober, C. P., Dana, J., & Rouder, J. N. (2018). Estimation accuracy in the psychological sciences. *PLoS One*, *13*(11), Article e0207239. <https://doi.org/10.1371/journal.pone.0207239>
- Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology*, *45*(2), 265–282. <https://doi.org/10.1111/j.2044-8317.1992.tb00992.x>
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360. <https://doi.org/10.1198/016214501753382273>
- Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, *20*(1), 101–148.
- Flack, V. F., & Chang, P. C. (1987). Frequency of selecting noise variables in subset regression analysis: A simulation study. *American Statistician*, *41*(1), 84–86. <https://doi.org/10.1080/00031305.1987.10475450>
- Freijeiro-González, L., Febrero-Bande, M., & González-Manteiga, W. (2022). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. *International Statistical Review*, *90*(1), 118–145. <https://doi.org/10.1111/insr.12469>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, *21*(2), 215–223. <https://doi.org/10.1080/00401706.1979.10489751>
- Gotlib, I. H., Borchers, L. R., Chahal, R., Gifuni, A. J., Teresi, G. I., & Ho, T. C. (2021). Early life stress predicts depressive symptoms in adolescents during the COVID-19 pandemic: The mediating

- role of perceived stress. *Frontiers in Psychology*, 11, Article 603748.  
<https://doi.org/10.3389/fpsyg.2020.603748>
- Greenwood, C. J., Youssef, G. J., Letcher, P., Macdonald, J. A., Hagg, L. J., Sanson, A., McIntosh, J., Hutchinson, D. M., Toumbourou, J. W., Fuller-Tyszkiewicz, M., & Olsson, C. A. (2020). A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One*, 15(11), Article e0242730. <https://doi.org/10.1371/journal.pone.0242730>
- Guo, P., Zeng, F., Hu, X., Zhang, D., Zhu, S., Deng, Y., & Hao, Y. (2015). Improved variable selection algorithm using a LASSO-type penalty, with an application to assessing Hepatitis B infection relevant factors in community residents. *PLoS One*, 10(7), Article e0134151.  
<https://doi.org/10.1371/journal.pone.0134151>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise or LASSO? analysis and recommendations based on extensive comparisons. *Statistical Science*, 35(4), 579–592.  
<https://doi.org/10.1214/19-STS733>
- Haynos, A. F., Wang, S. B., Lipson, S., Peterson, C. B., Mitchell, J. E., Halmi, K. A., Agras, W. S., & Crow, S. J. (2021). Machine learning enhances prediction of illness course: A longitudinal study in eating disorders. *Psychological Medicine*, 51(8), 1392–1402.  
<https://doi.org/10.1017/S0033291720000227>
- Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection—A review and recommendations for the practicing statistician. *Biometrical Journal. Biometrische Zeitschrift*, 60(3), 431–449.  
<https://doi.org/10.1002/bimj.201700067>
- Helwig, N. E. (2017). Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *Quantitative Methods for Psychology*, 13(1), 1–19.  
<https://doi.org/10.20982/tqmp.13.1.p001>
- Higham, N. J. (2002). Computing the nearest correlation matrix—A problem from finance. *IMA Journal of Numerical Analysis*, 22(3), 329–343. <https://doi.org/10.1093/imanum/22.3.329>
- IBM Corp. (2020). *IBM SPSS statistics for Windows* (Version 27.0) [Computer software]. IBM Corp.
- Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. *Advances in Methods and Practices in Psychological Science*, 2(1), 55–76.  
<https://doi.org/10.1177/2515245919826527>
- Laws, K. R. (2016). Psychology, replication & beyond. *BMC Psychology*, 4, Article 30.  
<https://doi.org/10.1186/s40359-016-0135-2>
- Liu, S., & Rhemtulla, M. (2022). Treating random effects as observed versus latent predictors: The bias–variance tradeoff in small samples. *British Journal of Mathematical & Statistical Psychology*, 75(1), 158–181. <https://doi.org/10.1111/bmsp.12253>
- Lovell, M. C. (1983). Data mining. *Review of Economics and Statistics*, 65(1), 1–12.  
<https://doi.org/10.2307/1924403>

- McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, *50*(5), 471–484. <https://doi.org/10.1080/00273171.2015.1036965>
- Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, *52*(1), 374–393. <https://doi.org/10.1016/j.csda.2006.12.019>
- Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, *34*(3), 1436–1462. <https://doi.org/10.1214/009053606000000281>
- Morozova, O., Levina, O., Uusküla, A., & Heimer, R. (2015). Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia. *BMC Medical Research Methodology*, *15*(1), Article 71. <https://doi.org/10.1186/s12874-015-0066-2>
- Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Allyn and Bacon.
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*(4), 434–447. <https://doi.org/10.1037/1082-989X.8.4.434>
- Pavlou, M., Ambler, G., Seaman, S., De Iorio, M., & Omar, R. Z. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, *35*(7), 1159–1177. <https://doi.org/10.1002/sim.6782>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Sauerbrei, W. (1999). The use of resampling methods to simplify regression models in medical statistics. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *48*(3), 313–329. <https://doi.org/10.1111/1467-9876.00155>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Sirimongkolkasem, T., & Drikvandi, R. (2019). On regularisation methods for analysis of high dimensional data. *Annals of Data Science*, *6*(4), 737–763. <https://doi.org/10.1007/s40745-019-00209-4>
- Smith, D. M., Wang, S. B., Carter, M. L., Fox, K. R., & Hooley, J. M. (2020). Longitudinal predictors of self-injurious thoughts and behaviors in sexual and gender minority adolescents. *Journal of Abnormal Psychology*, *129*(1), 114–121. <https://doi.org/10.1037/abn0000483>
- Su, W., Bogdan, M., & Candès, E. (2017). False discoveries occur early on the lasso path. *Annals of Statistics*, *45*(5), 2133–2150. <https://doi.org/10.1214/16-AOS1521>
- Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, *55*(4), 525–534. <https://doi.org/10.1177/0013164495055004001>

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1), 267–288.  
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tibshirani, R. J., & Taylor, J. (2012). Degrees of freedom in lasso problems. *Annals of Statistics*, 40(2), 1198–1232. <https://doi.org/10.1214/12-AOS1003>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Wester, R. A., Rubel, J., & Mayer, A. (2022). Covariate selection for estimating individual treatment effects in psychotherapy research: A simulation study and empirical example. *Clinical Psychological Science*, 10(5), 920–940. <https://doi.org/10.1177/21677026211071043>
- Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin*, 86(1), 168–174. <https://doi.org/10.1037/0033-2909.86.1.168>
- Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, 54(5), 719–750.  
<https://doi.org/10.1080/00273171.2019.1575716>
- Williams, D. R., & Rodriguez, J. E. (2020, March 3). *Why overfitting is not (usually) a problem in partial correlation networks*. PsyArXiv. <https://doi.org/10.31234/osf.io/8pr9b>
- Xu, C. J., van der Schaaf, A., Schilstra, C., Langendijk, J. A., & van't Veld, A. A. (2012). Impact of statistical learning methods on the predictive power of multivariate normal tissue complication probability models. *International Journal of Radiation Oncology\*Biophysics\* Physics*, 82(4), e677–e684. <https://doi.org/10.1016/j.ijrobp.2011.09.036>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122.  
<https://doi.org/10.1177/1745691617693393>
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 68(1), 49–67.  
<https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541–2563.
- Zhou, D. J., Chahal, R., Gotlib, I. H., & Liu, S. (2024). *Comparison of Lasso and stepwise regression in psychological data* [OSF project page containing literature review table, additional simulation results table, empirical data details table, R codes for generating simulation data of study]. OSF.  
[https://osf.io/uws2j/?view\\_only=9c0e8fbe8a8341d487598b7dc528fa0d](https://osf.io/uws2j/?view_only=9c0e8fbe8a8341d487598b7dc528fa0d)
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 67(2), 301–320.  
<https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Zou, H., Hastie, T., & Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *Annals of Statistics*, 35(5), 2173–2192. <https://doi.org/10.1214/009053607000000127>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429. <https://doi.org/10.1198/016214506000000735>



*Methodology* is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.