Lasso Regression: From Interpretation to Prediction

Authors: Zhang Lijin, Wei Xiayan, Lu Jiaqi, Pan Junhao, Pan Junhao

Date: 2020-05-14T00:00:00+00:00

Abstract

Traditional least squares regression methods emphasize accurate estimation of the current dataset, which can easily lead to model overfitting and compromise the reproducibility of model conclusions. With advances in the methodological domain, emerging statistical tools have emerged to address the limitations of traditional methods, and shifting from an excessive focus on interpreting regression coefficient values toward enhancing the predictive power of research results has become an increasingly important development trend in the field of psychology. The Lasso method, by introducing a penalty term into model estimation, can achieve higher predictive accuracy and model generalizability, while also effectively addressing overfitting and multicollinearity issues, thereby facilitating the construction and refinement of psychological theories.

Full Text

Lasso Regression: From Explanation to Prediction

ZHANG Lijin[1], WEI Xiayan[2], LU Jiaqi[2], PAN Junhao[1]* [1] Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China [2] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310028, China

Abstract

Traditional least squares regression focuses on accurate estimation of current datasets, which easily leads to model overfitting and undermines the replicability of findings. With methodological advances, emerging statistical tools can compensate for these limitations, and shifting from an excessive focus on interpreting regression coefficients to improving predictive power has become an important trend in psychology. By introducing penalty terms into model estimation, the Lasso method can achieve higher prediction accuracy and model generalizability

while effectively addressing overfitting and multicollinearity problems, thereby contributing to psychological theory construction and refinement.

Keywords: regression; regularization; prediction; Lasso

1. Introduction

The purpose of psychological research is to "describe, explain, predict, and influence behavior" (Peng & Li, 2011; Lippke & Ziegelmann, 2010), and examining relationships between variables is essential to achieving this goal. Regression analysis, as a method for evaluating variable relationships, has gained widespread acceptance and is available in all mainstream statistical software packages.

Regression analysis represents the most fundamental and classic quantitative method in social sciences (Xie, 2010), with many common statistical tests (e.g., ANOVA) being special cases of linear regression models. The general regression model can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} + \cdots + \beta_p x_{ip} + \varepsilon_i \quad (1)$$

This model contains p predictor variables, where β_0 is the intercept, β_j represents the regression coefficient for the jth predictor ($j = 1, 2 \ldots p$), y_i denotes the observed outcome for the ith participant, x_{ij} is the observation of the ith participant on the jth predictor, and ε_i is the residual term.

Regression analysis is commonly used to explore relationships between variables and to predict outcome variables. In psychological research employing regression models, Ordinary Least Squares (OLS) is the most frequently used method for estimating model coefficients (Helwig, 2017). OLS estimates parameters by minimizing the error between predicted and observed values of the outcome variable, providing the most accurate linear unbiased estimation for the current sample (Chartterjee, Hadi, & Price, 2000; Chartterjee & Hadi, 2006; Fomby, Hill, & Johnson, 1984; Maddala, 2002).

However, OLS' s focus on unbiased estimation of the current dataset readily leads to overfitting (Yarkoni & Westfall, 2017), where regression models based on the current sample perform poorly when fitting other samples from the same population or predicting future observations. This problem becomes more severe when many predictors exist, variables exhibit high collinearity, or data have low signal-to-noise ratios (Babyak, 2004; Helwig, 2017; McNeish, 2015). Overfitted models often include unnecessary redundant variables and overestimate the effects of some predictors, weakening model parsimony (Babyak, 2004; Cohen, J., Cohen, P., West, & Aiken, 2003; Derksen & Keselman, 1992). These issues significantly impact the generalizability and predictive utility of model conclusions.

With the flourishing development of machine learning, an increasing number of statistical tools have emerged to compensate for traditional methods' limitations. Regularization methods, exemplified by Lasso (Least absolute shrinkage and selection operator; Tibshirani, 1996), can effectively optimize OLS estimation and address overfitting problems (Candes & Tao, 2007; Tibshirani, 1996; Tibshirani, Saunders, Rosset, Zhu, & Knight, 2005; Zou, 2006; Zou & Hastie, 2005). Regularization methods introduce penalty terms into model estimation to shrink small regression coefficients to zero, sacrificing some estimation bias to achieve higher model prediction accuracy and generalizability. This approach can compress redundant predictors' coefficients to zero, simultaneously performing variable selection and effectively avoiding generalizability problems caused by overfitting, yielding more parsimonious and efficient predictive models that contribute to psychological theory construction and refinement.

Since its introduction, Lasso has attracted considerable research attention (Zou, Hastie, & Tibshirani, 2007). Due to its excellent performance in variable selection and model stability, many researchers in medicine, economics, neuroscience, and other fields have adopted Lasso for predictive modeling (e.g., Fontanarosa & Dai, 2011; Lee, Chao, Ting, Chang, Huang, Wu, et al., 2014; Nguyen, Duong, Venkatesh, & Phung, 2015). However, outside neuroscience, Lasso has been rarely used in psychology (Johnson & Sinharay, 2011; McNeish, 2015; Yarkoni & Westfall, 2017). The main obstacle stems from 质疑 about the interpretability of machine learning methods like regularization, which often don' t rely on traditional hypothesis testing and adopt a more data-driven approach for exploration and prediction, thus being considered a "black box." Wu (2019) pointed out that individual regression coefficients in regression models are similarly uninterpretable. For example, regression results are often reported as: "When holding other predictors constant, a one-unit change in this predictor leads to a β-unit change in the outcome," yet this condition is almost impossible to satisfy. Beyond interpreting research conclusions, model generalizability and predictive ability also warrant attention.

In psychological research, limited by computational power and traditional statistical methods, researchers have primarily used hypothesis testing to verify theories and examine variable relationships. As these methods have become widespread, their limitations have become increasingly prominent, with overfitting and the replicability crisis receiving growing attention (Hu et al., 2016; Nuzzo, 2014). With machine learning' s rapid development, emerging data science tools have demonstrated tremendous value in healthcare and other fields, and improving predictive power will become an important future trend in psychology (Yarkoni & Westfall, 2017).

This paper uses Lasso as an example to comprehensively introduce its principles, implementation steps, and advantages to psychological researchers, combining theoretical foundations with practical analysis and application status. We call on researchers to adopt more robust Lasso regression to improve conclusion generalizability, especially with small samples or many variables. Additionally, this paper introduces various Lasso extensions and their applications in network analysis and latent variable modeling, hoping to provide references for practical applications and encourage more psychological researchers to embrace these emerging data science tools.

2. Traditional Methods and Their Limitations

In standard OLS regression, parameter estimation is obtained by minimizing the loss function—the vertical squared distance between observed and predicted values. The OLS loss function is specifically formulated as (McNeish, 2015):

$$L_{OLS}(\beta) = \|Y - X\beta\|^2 \quad (2)$$

where L_{OLS} is the loss function, assuming n observations and p predictors (including intercept), $X(n \times p)$ and $Y(n \times 1)$ are the predictor matrix and outcome vector respectively, and $\beta(p \times 1)$ is the vector of regression coefficients.

By minimizing L_{OLS}, OLS regression yields the Best Linear Unbiased Estimator (BLUE) $\hat{\beta}_{OLS}$, with low computational burden suitable for many modeling situations in psychology. However, when studies include many predictors, OLS estimation has several limitations:

First, **overfitting**: The regression model becomes overly complex, where some parameters' significance results from sampling variability, making the model applicable only to the current sample and lacking generalizability.

Prediction error can be decomposed into bias and variance components. Bias refers to the difference between predicted and true values, while variance refers to the dispersion of predictions. OLS estimation aims to reduce prediction error by controlling estimation bias, but this increases between-sample variance of parameters. Current parameter estimates may apply only to the current dataset and are susceptible to minor fluctuations across samples (as shown in Figure 1a [Figure 1: see original paper], where although the model fits data points accurately with small bias, it may not apply to other samples), leading to overfitting. Overfitting causes models to overestimate regression coefficients while underestimating their standard errors, easily revealing significant predictive effects for unrelated redundant variables. Results may apply only to the current sample rather than the population. The risk of misinterpreting parameters increases as the ratio of observations (n) to predictors (p) decreases (i.e., insufficient sample size) (Babyak, 2004; Derksen & Keselman, 1992).

Conversely, if parameter estimates from the current dataset have acceptable bias, between-sample variance decreases due to biased estimation, yielding more generalizable results (as shown in Figure 1b). Therefore, in practical data analysis, we must properly address this bias-variance tradeoff problem. Traditional OLS estimation focuses on precise estimation of the current dataset, inevitably leading to overfitting results when many predictors are included, thereby weakening model generalizability.

[Figure 1: see original paper] Figure 1. Bias-Variance Tradeoff

Second, **multicollinearity**: When multiple predictors in a regression model are correlated, with complete multicollinearity occurring when predictor correlations equal ±1. With strong multicollinearity, OLS-estimated regression coefficients become highly sensitive to minor fluctuations in sample data, yielding poor estimation stability. Regression coefficient estimation variance also increases with stronger collinearity among independent variables (Zhang, 2010). That is, when replacing some data in the sample, regression coefficients change substantially due to multicollinearity. This not only results in models lacking generalizability but can also make some important variables' regression coefficients trivial or even opposite to reality (Rao, 1976).

Additionally, when models contain many predictors, researchers often use stepwise regression to add or delete variables to obtain an effective predictor set. However, this method violates the prerequisite assumption of regression inference that all predictors exist as a fixed whole (Lockhart, Taylor, Tibshirani, R. J., & Tibshirani, R., 2014). Overfitting problems become more pronounced when using stepwise regression for model selection. At this point, t-tests or F-tests for statistical inference not only fail to follow their appropriate null hypothesis distributions but also lack appropriate degrees of freedom for analysis. Basic statistical tests and associated p-values become unsuitable for model selection with continuously changing variables. This model selection may increase the Type I error rate for regression coefficient hypothesis testing (Wilkinson, 1979).

3. The Lasso Method

3.1 Introduction to Lasso

Compared with OLS estimation mentioned above, regularization methods introduce penalty functions to the OLS loss function to penalize overly complex models. The specific formula can be expressed as:

$$L_{Reg}(\beta) = L_{OLS}(\beta) + \lambda P(\beta) \quad (4)$$

where $L_{Reg}(\beta)$ is the penalized loss function, $L_{OLS}(\beta)$ represents the standard OLS loss function, $P(\beta)$ represents the penalty function, and $\lambda(\geq 0)$ represents the tuning parameter controlling the degree of coefficient shrinkage, with larger values indicating stronger penalty. When $\lambda = 0$, the loss function imposes no penalty, and $L_{Reg}(\beta)$ becomes the OLS loss function. Different penalty functions $P(\beta)$ correspond to different regularization methods.

Lasso, as one regularization method, uses the sum of absolute values of regression coefficients as its penalty function to shrink coefficients, i.e., $P_{Lasso}(\beta) =$

$\lambda \sum_{j=1}^{p} |\beta_j|$. In parameter estimation, since absolute value signs are difficult to decompose operationally, $|\beta_j|$ can be converted to $\pm 1 \times \beta_j$, where the sign (+1 or -1) matches β_j' s sign. Thus, the Lasso loss function formula can be expressed as (McNeish, 2015):

$$L_{Lasso}(\beta) = (|Y - X\beta|)^2 + \lambda W^T \beta \quad (5)$$

In this formula, L_{Lasso} refers to the Lasso regression model' s loss function, $X(n \times p)$, $Y(n \times 1)$, and $\beta(p \times 1)$ are the predictor matrix, outcome vector, and coefficient vector respectively, while $W(p \times 1)$ is a vector with values ± 1 (signs matching corresponding values in the β vector).

Compared with other regularization methods (e.g., Ridge regularization uses the sum of squared coefficients as its penalty function, applying less shrinkage to small coefficient estimates, making it difficult to compress redundant predictors' coefficients to zero, and more likely to over-shrink important coefficients (Hesterberg, Choi, Meier, & Fraley, 2008)), Lasso can directly compress redundant predictors' coefficients to zero, serving a variable selection function to obtain a streamlined and more efficient predictor set (Tibshirani, 1996), while also reducing over-shrinkage of important coefficients.

Yarkoni and Westfall (2017) noted that compared with OLS estimation, models obtained via Lasso typically generalize better to new datasets. In OLS regression, model R^2 (i.e., explained variance in the outcome) typically increases with model complexity. Lasso, however, focuses not only on explaining the current dataset (i.e., obtaining higher R^2) but also on obtaining more parsimonious models that better generalize to the population. This shift from explanation to prediction directs research not just toward the past (interpreting the current dataset) but also toward the future (predictive ability for new datasets). This characteristic not only contributes to psychological theory construction and refinement but also helps reduce the impact of the replicability crisis.

Furthermore, Lasso avoids overfitting and multicollinearity problems associated with OLS estimation when many predictors are present. Theoretical inadequacy and predictor collinearity are common phenomena in psychology. When researchers' theoretical assumptions are unclear, using confirmatory methods with multiple testing corrections (like stepwise regression) is theoretically wrong (Serang, Jacobucci, Brimhall, & Grimm, 2017), as later-included variables are often weakened due to correlation with previously included variables (Frank & Heiser, 2011). Lasso treats the predictor set as a whole, better addressing this issue.

Due to the penalty term' s introduction, Lasso requires relatively higher computational effort. The Least Angle Regression (LARS) estimation method proposed by Efron, Hastie, Johnstone, and Tibshirani (2004) to address this issue is now widely applied. For applied researchers, as regularization methods have matured, R software packages for direct Lasso regression modeling have been developed, which this paper will detail below.

3.2 Lasso Regression Implementation Steps

Lasso regression modeling typically includes two parts: selection of parameter λ and calculation of p-values. The following sections detail the methodological principles, while the appendix demonstrates Lasso regression implementation in R with detailed comparisons to OLS regression.

3.2.1 Parameter λ Selection Parameter λ selection determines the degree of coefficient shrinkage, with different λ values potentially producing different results. Currently, two common methods exist for selecting the optimal λ value (McNeish, 2015):

The first method is cross-validation from machine learning. The process is as follows: First, data are divided into K equally sized samples, typically with K being 5, 10, or N (sample size). Then, for a chosen λ value, Lasso estimation is applied to $K-1$ samples, and the resulting coefficients are used to validate the Kth sample, testing model adequacy. This process is repeated K times. Finally, we obtain model fit values (e.g., mean squared error) and standard errors for a given λ. Cross-validation typically repeats this process for 100 different λ values, selecting λ based on mean standard error magnitude. Generally, we select the λ value with minimum mean standard error, though this sometimes means insufficient coefficient shrinkage may not fully solve overfitting. Therefore, some research suggests selecting the λ value corresponding to one standard error above the minimum mean standard error (Waldmann, Mészáros, Gredler, Fuerst, & Sölkner, 2013).

The alternative method is information criteria. The λ selection process is similar to cross-validation: for multiple λ values, Lasso models are fitted (using all data) at each value, and information criterion values (e.g., Akaike Information Criterion, AIC; Bayesian Information Criterion, BIC) are calculated. The specific formulas are:

$$AIC = n\log(RSS/n) + 2df \quad (6)$$

$$BIC = n\log(RSS/n) + \log(n)df \quad (7)$$

where RSS refers to residual sum of squares and df refers to degrees of freedom. Typically, we select the λ value producing a local or global minimum information criterion (McNeish, 2015).

Currently, most researchers primarily use cross-validation to determine λ values (Obuchi & Kabashima, 2016).

3.2.2 p-value Calculation Most variable selection methods (e.g., stepwise regression) yield incorrect degrees of freedom or standard errors. These methods test not the k predictors that should exist as a whole, but the m predictors selected after screening ($m \leq k$, Thompson, 2001). For example, for a regression model with sample size $n = 101$ and $k = 50$ predictors, the F-test degrees of freedom should be (50, 50), but if stepwise regression selects five predictors from 50, the F-test degrees of freedom become (5, 95). Consequently, calculated p-values are often unreliable (Lockhart et al., 2014). Currently, no good methods exist for handling p-value calculation without repeated sampling or data splitting.

In Lasso regression, calculating standard errors and significance for coefficients not compressed to zero is also difficult. To address this, Lockhart et al. (2014) proposed a method for calculating p-values in Lasso estimation without repeated sampling or data splitting. This method resembles traditional likelihood ratio tests. In standard likelihood ratio tests, we calculate deviance (deviance = -2log(likelihood)) for full and restricted models (where some freely estimated parameters in the full model are constrained to zero in the restricted model), then use chi-square tests to compare nested models (restricted nested within full) for model selection. Similarly, Lockhart et al. (2014) demonstrated that covariance between observed outcome values (Y) and model predictions ($X\hat{\beta}$) can serve a "deviance-like" function. That is, by calculating the change in model covariance when adding a predictor (from a restricted model where that predictor' s coefficient is constrained to zero to a full model where all coefficients are freely estimated), significance testing can achieve variable selection. This method incorporates effects of all other predictors when testing each predictor' s significance, avoiding the influence of earlier-included variables on later-included variables in stepwise regression, and doesn' t require data splitting or repeated sampling for inferential testing, making it relatively straightforward.

To demonstrate Lasso regression implementation steps and reporting standards, the appendix provides a detailed demonstration using empirical data, comparing Lasso and OLS regression methods. The analysis uses the glmnet package (Friedman, Hastie, & Tibshirani, 2010) for λ selection and the covTest package (Lockhart et al., 2014) for p-value calculation.

4. Applications of Lasso Regression

Lasso regression's advantages primarily manifest as a stable variable filter and in building models with greater generalizability and predictive power. In research with relatively underdeveloped theory, researchers need such methods to avoid over-interpreting current samples and explore patterns applicable to populations. This shift from explanation to prediction helps enhance theoretical significance and applied value of such research.

Lasso' s excellent properties give it broad application prospects in educational psychology, clinical psychology, developmental psychology, and other fields.

However, only a few psychological studies have adopted Lasso methods (e.g., Hartmann, Zeeck, & Barrett, 2010; Scheidt et al., 2012; Schmid, Taylor, Foldi, Berres, & Monsch, 2013). McNeish (2015) also noted a substantial disconnect between the application status of statistical methods in psychology and advances in statistical research. The translation of statistical research findings to broad psychological application often takes considerable time, preventing applied fields from quickly benefiting from the latest statistical research. Therefore, the following sections illustrate Lasso' s practical applications in clinical psychology and cognitive neuroscience to demonstrate its specific use and advantages, hoping to provide references for researchers.

4.1 Lasso Applications in Cognitive Neuroscience

In neuroscience, Lasso has been successfully applied in genome-wide association studies (GWAS) or candidate gene studies to screen single nucleotide polymorphisms (SNPs; Ayers & Cordell, 2010; Shi et al., 2011), detect gene-gene interactions (D' Angelo, Rao, & Gu, 2009; Li, Das, Fu, Li, R., & Wu, 2011), and conduct risk prediction based on GWAS results (Kooperberg, LeBlanc, & Obenchain, 2010). GWAS can identify risk genes affecting neurological and psychiatric disorders, but such research often involves numerous genetic loci and faces replicability challenges (Kohannim et al., 2012). Using Lasso can appropriately reduce the number of SNPs, screen for genes stably associated with outcome variables, and build replicable models. Additionally, traditional GWAS analysis treats each gene' s effect as independent, ignoring potential linkage disequilibrium (LD) structures where some variants are more likely to be inherited together. In summary, Lasso' s advantages in genetic analysis include (Cho, Kim, Oh, Kim, & Park, 2009; Cho et al., 2010; Lin et al., 2009; Malo, Libiger, & Schork, 2008; Shi et al., 2011): (1) handling the multidimensional nature of genomes; (2) addressing multicollinearity caused by LD; and (3) handling multiple comparison problems.

Kohannim et al. (2012) used Lasso regression to reduce the number of correlated genes and screen for genes reliably associated with brain structure. They collected whole-genome data and relevant covariates from 729 elderly participants, with temporal lobe volume (a biomarker for neurodegenerative disease) as the outcome variable. Lasso regression identified the most effective set of SNPs affecting the outcome from candidate SNPs, ultimately obtaining 22 genes that significantly affected temporal lobe volume. To test the genetic results'replicability, they conducted validation of the most relevant MACROD2 gene in an independent sample of healthy young adults, finding consistent effects of MACROD2 on brain structure, thereby validating the robustness of Lasso-derived genetic associations.

4.2 Lasso Applications in Clinical Psychology

Due to difficulties in collecting clinical samples and incomplete understanding of many psychological disorders, clinical research often considers numerous variables, resulting in low ratios of observations to predictors (Demjaha et al., 2017). Additionally, clinical assessment requires building models for stable inference. Traditional stepwise regression for variable selection easily leads to overfitting, whereas Lasso can obtain stable parameter estimates and improve prediction accuracy (Harrell, 2015), better meeting clinical assessment requirements.

Accordingly, Demjaha et al. (2017) investigated factors affecting treatment resistance in first-episode psychosis, tracking 323 patients and using Lasso multiple regression to analyze relationships between treatment resistance and clinical/demographic variables. Lasso multiple regression results showed that schizophrenia diagnosis, negative symptoms, younger age at first onset, and longer duration without psychiatric treatment significantly predicted treatment resistance.

Furthermore, early patient identification is a prerequisite for effective clinical intervention and treatment. Lasso has been successfully applied to identify potential patients. Schmid, Taylor, Foldi, Berres, and Monsch (2013) conducted an eight-year follow-up of 29 patients who later developed Alzheimer' s disease and 29 matched normal controls, examining objective behavioral measures and neuropsychological functional changes. With research variables ($k = 115$) far exceeding observations ($n = 29$), ordinary regression methods would cause severe overfitting. To obtain more predictive models, researchers used Lasso regression to identify which variables could early distinguish those who would develop Alzheimer' s disease from normal controls. Ultimately, 11 most predictive variables were selected from 115 predictors, effectively distinguishing the two groups early.

5. Lasso Extensions

5.1 Extended Forms of Lasso

Building upon Lasso, researchers have developed various regularization models using different penalty functions based on predictor characteristics, including Relaxed Lasso (Meinshausen, 2007), Adaptive Lasso (Zou, 2006), Bayesian Lasso (Park & Casella, 2008), Fused Lasso (Tibshirani et al., 2005), and Group Lasso (Yuan & Lin, 2006). The following sections introduce principles and corresponding R packages for several Lasso extensions.

5.1.1 Relaxed Lasso When the number of observed indicators p far exceeds sample size N, Lasso converges slowly (Fan & Peng, 2004). Since Lasso cannot simultaneously achieve satisfactory compromise between computational complexity and convergence speed, Meinshausen (2007) proposed a two-stage analysis method—Relaxed Lasso. In Relaxed Lasso, model selection and parameter estimation are separated into two independent processes. The method first uses ordinary Lasso regression to screen appropriate predictors, then conducts coefficient estimation on selected variables. An adjustment parameter Φ changes penalty strength ($\lambda_2 = \Phi \times \lambda$, $1 > \Phi \geq 0$, where λ and λ_2 are tuning parameters in the first and second stages respectively), weakening or eliminating penalty effects to reduce coefficient estimation bias. When $\Phi = 1$, coefficient estimates match those from ordinary Lasso; when $\Phi = 0$, coefficient estimates match OLS estimates. Relaxed Lasso achieves faster convergence than Lasso while maintaining computational complexity (Meinshausen, 2007). Theoretical and numerical results show that for high-dimensional data, Relaxed Lasso produces sparser models with equal or smaller prediction loss than Lasso.

For Relaxed Lasso applications, complete software packages are available. The R package relaxo (Meinshausen, 2019) specializes in Relaxed Lasso analysis, conveniently obtaining solutions by simply calling cvrelaxo or relaxo functions. This paper also demonstrates Relaxed Lasso regression using empirical data (with Φ fixed at 0 in the second stage, i.e., using OLS regression), finding that Relaxed Lasso achieves prediction power similar to OLS using five variables with only two predictors.

5.1.2 Adaptive Lasso Lasso controls coefficient shrinkage through tuning parameter λ (Tibshirani, 1996). When researchers select and fix λ at a specific value via cross-validation, Lasso applies equal penalty to all variables. Although this reduces over-shrinkage of important coefficients compared to Ridge regularization, it may still compress important variables' coefficients, creating estimation bias (Fan & Li, 2001). Zou (2006) improved Lasso by adding adaptive weights before the penalty term, proposing Adaptive Lasso. In Adaptive Lasso, for any chosen $\gamma > 0$, the weight vector $\hat{w} = 1/|\hat{\beta}|^{\gamma}$, where OLS estimates can serve as initial coefficient estimates $\hat{\beta}$. The Adaptive Lasso penalty term becomes:

$$P_{aLasso}(\beta) = \lambda \sum_{j=1}^{p} \hat{w}_j |\beta_j|$$

Adaptive Lasso' s adaptive weights depend on data, with different variables' coefficients receiving different penalty degrees. Variables with larger initial coefficient estimates receive smaller weights and thus less penalty, while variables with smaller initial estimates receive larger weights and greater penalty. Therefore, Adaptive Lasso variable selection makes important variables more likely to enter the model while unimportant variables are more easily eliminated, better achieving variable selection while reducing coefficient estimation bias. Compared to Lasso, Adaptive Lasso is more suitable when the ratio of observed indicators p to sample size N is very large. Currently, R packages including glmnet (Tibshirani et al, 2019), msgps (Hirose, 2019), and parcor (Kraemer & Schaefer, 2019) can perform Adaptive Lasso analysis. Additionally, SAS software' s Proc GlmSelect can implement Adaptive Lasso.

5.1.3 Bayesian Lasso In the frequentist framework, Lasso achieves regularization by adding penalty terms to the likelihood function. In Bayesian methods, appropriate prior distributions can play the role of penalty terms through their log forms. For example, Tibshirani (1996) noted that in Bayesian methods, providing independent double-exponential prior distributions $\lambda \exp(-\lambda|\theta_j|)$ for parameters θ_j can achieve Lasso regularization. Double-exponential prior distributions are unimodal and symmetric like zero-mean normal distributions but have greater kurtosis. Larger λ values concentrate the probability density function more around zero.

Furthermore, frequentist algorithms implementing Lasso (e.g., Efron et al., 2004; Friedman et al., 2010; Wu & Lange, 2008) cannot provide valid standard error estimates, hindering Lasso applications in frequentist domains (Kyung, Gill, Ghosh, & Casella, 2010). Bayesian Lasso can provide valid standard error estimates via Gibbs sampling (Kyung et al., 2010). Bayesian Lasso regression models proposed by Park and Casella (2008) and Hans (2009) can estimate both unknown coefficients and regularization parameters simultaneously, avoiding substantial computational burden from traditional cross-validation and showing broad application prospects. Applied researchers can conveniently conduct Bayesian Lasso regression modeling using the R package blasso (Gramacy, 2019).

5.2 Extended Applications of Lasso

Beyond regression models, Lasso can be used to screen mediators (Serang et al., 2017), and regularization methods are increasingly applied to structural equation modeling and psychological network models (Epskamp, Borsboom, & Fried, 2018).

5.2.1 Latent Variable Models Latent variable models are primarily used to analyze questionnaire data, considering measurement error in model estimation. In latent variable modeling, regularization methods have attracted methodologists' attention and been gradually introduced into structural equation modeling, such as using Bayesian Ridge or Lasso regularization to address overly restrictive traditional confirmatory factor analysis (Muthén & Asparouhov, 2012; Pan, Ip, & Dubé, 2017), and using regularization methods for predictor selection in MIMIC models (Multiple Indicators and Multiple Causes; Jacobucci, Brandmaier, & Kievit, in press).

Currently, the most popular latent variable analysis software Mplus (Muthén, L, K., & Muthén, B, O., 1998-2019) can already conduct structural equation modeling using Ridge regularization, with widespread application (Zhang, Lu, Wei, & Pan, 2019). Specialized R packages "blcfa" (Pan, Zhang & Ip, 2019) for Bayesian Lasso confirmatory factor analysis and "regsem" (Jacobucci, 2019) are also available to help researchers use Ridge or Lasso regularization for exploratory factor analysis and MIMIC modeling. Unfortunately, due to Lasso ...

5.2.2 Network Models Psychological network models use nodes to represent observable variables and edges to represent connections between them, with edge weights representing connection strength. These models assume that certain psychological processes and states (e.g., cognitive processes, psychopathological symptoms) occur simultaneously, focusing on interactions among observable variables in the network. Psychological network models help researchers deeply understand relationships among observable variables and serve as powerful supplements to latent variable models. Recently, these models have been widely applied in personality psychology and clinical psychology research (e.g., Costantini et al., 2019; Richetin, Preti, Costantini, & De Panfilis, 2017).

Since these models examine many variables and parameters, researchers typically combine Lasso methods with network analysis for variable selection to avoid overfitting and reduce Type I error rates. Adaptive Lasso and Graphical Lasso can help researchers obtain sparse, more generalizable network models. For example, Marcus, Preszler, and Zeigler-Hill (2017) used Adaptive Lasso to build a Dark Personality network model; Costantini et al. (2015a) developed an implicit measurement tool for conscientiousness variables based on Adaptive Lasso network models; Di Pierro, Costantini, Benzi, Madeddu, and Preti (2018) used Graphical Lasso to build a psychopathological network model of narcissistic traits. Such network models can be implemented via qgraph (network analysis package; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012) and glasso (Graphical Lasso package; Friedman, Hastie, & Tibshirani, 2019). To facilitate applied researchers, Costantini et al. (2015b, 2019) detailed principles and R implementation methods for Adaptive Lasso and Graphical Lasso network models.

6. Discussion

6.1 Application Recommendations

In psychological research, researchers often focus primarily on interpreting variable relationships. However, Yarkoni and Westfall (2017) noted that this perspective has led to numerous psychological studies exploring complex psychological mechanisms whose models poorly predict future behavior. As replicability issues gain attention, using statistical methods and standardizing research processes to provide solutions to the replicability crisis has become a hot topic in psychology (Giordano & Waller, 2019; Hu et al., 2016; Spellman, 2015). Overfitting from excessive focus on interpreting current datasets is a key contributor to the replicability crisis, and researchers have proposed various countermeasures. For example, calculating sample size based on power and effect size before experiments, and lowering the p-value threshold to 0.005 while increasing sample size to reduce Type II error rates (Benjamin et al., 2018). However, some studies (e.g., clinical research) cannot collect sufficient sample sizes, and having many variables is common when theory is underdeveloped, making overfitting more severe (Babyak, 2004; McNeish, 2015).

Therefore, researchers have noted that new statistical tools (e.g., regularization methods, Bayesian methods) may avoid hypothesis testing limitations and reduce the replicability crisis (Hu et al., 2016; Benjamin et al., 2018). Increasingly, researchers suggest that machine learning tools may help psychology become a more predictive science, and the shift from explanation to prediction may help researchers better understand behavior and its underlying mechanisms (Rosenberg, Casey, & Holmes, 2018; Serang et al., 2017).

Regularization models represented by Lasso play increasingly important roles in machine learning and have been widely applied in biomedicine and other fields. In psychology, regularized sparse models can also help researchers conduct variable selection, solve overfitting problems, and control Type I error rates (Liu, Cui, Liu, & Luo, 2015; Xu, Wang, Sun, & Wang, 2017). Lasso shows superior performance with small samples and many variables and is increasingly applied in psychology. Beyond clinical psychology and cognitive neuroscience, Lasso regression can also be valuable in educational psychology and personality psychology. Therefore, this paper aims to demonstrate the value of regularization models through Lasso regression principles and applications, promoting greater use of machine learning tools in psychology. We also call on applied researchers to adopt Lasso methods when variable numbers are large or sample sizes are insufficient.

6.2 Limitations and Future Directions

The main obstacle to Lasso regression application is difficulty obtaining standard error estimates. This affects p-value calculation, though Lockhart et al.' s (2014) method and corresponding R package effectively address this issue. However, we also hope researchers can move beyond significance testing thinking when applying such machine learning methods and focus more on overall model predictive ability. On the other hand, inability to obtain standard errors also affects effect size and confidence interval calculation, but Bayesian Lasso can effectively estimate standard errors and credible intervals, compensating for this limitation. With Bayesian statistics gaining popularity (Van de Schoot et al., 2017), Bayesian Lasso is expected to see deeper development and application.

Additionally, many mainstream statistical software packages cannot implement Lasso regression (e.g., SPSS, Mplus), greatly hindering its application. While many R packages can implement Lasso, each has limitations. Hadley Wickham, Chief Scientist at RStudio and author of ggplot2, noted in an interview (Qiu, 2019) that he suggests students try more robust regression methods like Lasso in his classes, but pointed out that approximately 13 R packages exist for Lasso methods, each being imperfect (e.g., cannot handle missing values, categorical variables), and he plans to integrate these packages to create a more efficient analysis tool. We believe that as regularization models and their supporting analysis tools mature, applied researchers will more conveniently adopt regularization methods for modeling.

Finally, Lasso applications beyond regression models have just begun, and Lasso' s excellent properties make it potentially valuable for complex models (e.g., latent interaction models, intensive longitudinal models). We hope methodologists can fully leverage Lasso' s value across various domains as the method develops. Future research should also compare Lasso with other regularization methods and explore their respective applicable modeling scenarios to provide recommendations for applied research.

Appendix 1: Lasso Regression Demonstration

To verify that traditional OLS estimation easily leads to overfitting, demonstrate Lasso regression steps and reporting standards, and promote Lasso regression application, this paper uses a demonstration with detailed analysis procedures and comparisons to traditional estimation methods. The demonstration also includes Relaxed Lasso analysis. Analysis uses R software, with detailed code provided in Appendix 2.

Data come from 395 Portuguese middle school students (Cortez & Silva, 2008), containing 11 continuous variables: (1) age, (2) family relationship quality (famrel), (3) free time after school (freetime), (4) frequency of going out with friends (goout), (5) weekday alcohol consumption frequency (dalc), (6) weekend alcohol consumption frequency (walc), (7) self-rated health (health), (8) number of absences, (9) first math test score (G1), (10) midterm test score (G2), and (11) final exam score (G3). The final exam score serves as the outcome variable, investigating factors that effectively predict final math performance. Correlation analysis shows strong positive correlations between first test scores, midterm scores, and final exam scores.

[Figure 2: see original paper] Figure 2. Correlation matrix among variables
Note: Red shades represent negative correlations, blue shades represent positive correlations; darker colors indicate stronger correlations.

In Lasso regression, 10-fold cross-validation first selects the appropriate penalty term λ using R' s glmnet package (Friedman, Hastie, & Tibshirani, 2010). Notably, to ensure consistent λ results across cross-validation analyses, the set.seed() function must be used to set a random seed; otherwise, each run …

Results show the λ minimizing mean squared error (MSE) is 0.043, and $\lambda +$ 1se is 0.776. Figure 3 [Figure 3: see original paper] presents MSE changes as log(lambda) increases. As λ' s penalty on complex models increases, MSE also increases, and increasing penalty eventually compresses all coefficients to zero, maximizing MSE.

[Figure 3: see original paper] Figure 3. Ten-fold cross-validation results
Note: The two vertical lines represent λ values minimizing MSE and $\lambda + 1se$.

Figure 4 [Figure 4: see original paper] shows standardized regression coefficient compression as log(lambda) increases. As penalty strength increases, all standardized coefficients are eventually compressed to zero. At $\lambda = 0.776$, two coefficients remain non-zero. According to output results, two predictors are retained: G1 (first math test score) and G2 (midterm test score).

[Figure 4: see original paper] Figure 4. Coefficient shrinkage by penalty term

Additionally, the covTest package (Lockhart et al., 2014) can calculate p-values in Lasso regression. Further p-value calculation shows only G1 and G2 variables pass significance testing (Table 1).

In contrast, OLS estimation identified five variables as significantly predicting final math scores: age, family relationship quality, absences, first test score, and midterm score (Table 1). However, results show absences positively predicted final scores—more absences predicted higher final scores ($b = 0.042, p = 0.001$) —which clearly contradicts common sense. Correlation analysis also found no significant correlation between absences and final scores ($r = 0.034, p = 0.497$). OLS regression' s significant results likely occurred because the low ratio of sample size to predictors ($n/p = 3.95$) caused overfitting, where the model incorrectly learned non-existent patterns while minimizing differences between predicted and observed outcome values. Moreover, compared to Lasso regression, OLS' s two additional significant predictors had weak correlations with final scores (Figure 2). Age showed significant negative correlation with final math scores ($r = -0.162, p = 0.001$), while family relationship quality showed no significant correlation ($r = 0.051, p = 0.309$).

Further Relaxed Lasso analysis used OLS regression to model final math scores with G1 and G2 selected by Lasso. Results showed Relaxed Lasso's R^2, adjusted R^2, and mean squared error were similar to traditional OLS. Thus, Relaxed Lasso achieved prediction power comparable to OLS using five variables with only two predictors.

Table 1. Lasso, OLS, and Relaxed Lasso regression results

Coefficient estimates (p-value)	Lasso	OLS	Relaxed Lasso
famrel	-0.206(0.009)**	0.36(0.001)**	
freetime	0.058(0.57)	-0.014(0.891)	-0.108(0.448)
goout	0.17(0.105)		
health	0.046(0.509)		
absences	0.042(0.001)**		
G1	0.164(0.003)**	0.977(<0.001)***	0.153(0.007)**
G2	0.057(0.005)**	0.903(<0.001)***	0.987(<0.001)***
R^2		0.987(<0.001)***	
Adjusted R^2			
Mean Square Error			

Note: ***indicates*** $p < 0.01$, indicates $p < 0.001$.*

These analyses demonstrate that OLS regression' s selected predictors may be unreliable and redundant. On one hand, OLS-selected predictors in this study had weak correlations with the outcome. On the other hand, the three additional predictors didn' t substantially improve outcome explanation, with R^2 and adjusted R^2 values similar to the two-predictor model. Moreover, Relaxed Lasso avoids Lasso' s simultaneous shrinkage of non-zero coefficients (G1, G2) while compressing unimportant coefficients.

Notably, Lasso regression doesn't always yield more parsimonious predictor sets; its purpose is to achieve higher predictive power with fewer predictors. This is particularly evident with small samples, where OLS hypothesis testing typically yields higher standard error estimates and lower power to control Type I error rates, while Lasso more easily achieves higher power and predictive ability.

Appendix 2: Lasso Regression Example Code

```
student <- read.table("mat_2.txt", sep="\t", header=FALSE)
IV <- (student[,1:10])
IV1 <- scale(IV, FALSE, FALSE)  # No standardization of predictors

# 10-fold cross-validation
install.packages('glmnet')
library(glmnet)
set.seed(1222)  # Set random seed for consistent cross-validation results
Lambda <- cv.glmnet(IV1, student[,11])  # Lasso regression results
coef(Lambda, s=Lambda$lambda.1se)
abline(v=log(Lambda$lambda.min))
savePlot(filename = "loglambda", type ="png", device = dev.cur(), restoreConsole = TRUE)

# Calculate p-values using covTest package
library('devtools')
install_github('cran/covTest')  # covTest not available on CRAN, install from GitHub via de
library(covTest)
IV <- student[,1:10]
df <- nrow(IV) - 1
IV2 <- scale(IV, TRUE, TRUE)/sqrt(df)  # Standardize predictors
LarsCoef <- lars(IV2, student[,11])
covTest(LarsCoef, IV2, student[,11])  # Calculate p-values
```

Lasso regression: From explanation to prediction

ZHANG Lijin[1], WEI Xiayan[2], LU Jiaqi[2], PAN Junhao[1]
([1] Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China)
([2] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310028, China)

Abstract: Psychological research focuses on describing, explaining, and predicting behavior, and understanding associations between variables is essential to this process. Regression analysis, as a method for evaluating variable relationships, is widely used in psychological studies. However, due to its heavy focus on interpreting sample data, traditional ordinary least squares regression has several drawbacks, such as overfitting and limited capacity to handle multicollinearity, which may undermine model generalizability. These drawbacks inevitably influence the promotion and prediction of model conclusions.

With rapid methodological development, Least absolute shrinkage and selection operator (Lasso) regression has emerged to better compensate for traditional methods' limitations. By introducing a penalty term and shrinking regression coefficients to zero, Lasso regression can achieve higher model prediction accuracy and generalizability at the cost of some estimation bias. Additionally, Lasso regression can effectively address multicollinearity problems and has been widely used in medicine, economics, neuroscience, and other fields.

In psychology, due to limited computational power, researchers have traditionally relied on hypothesis testing to understand variable associations and verify theories. Now, with machine learning' s rapid development, a shift from focusing on coefficient interpretation to improving model prediction has emerged and become increasingly important. Based on fundamental theories and real data analysis, this paper aims to introduce Lasso regression principles, implementation steps, and advantages. With statistical science' s help, more applied researchers will hopefully focus on emerging statistical tools to promote psychology' s development.

Key words: regression; regularization; Lasso; prediction

Note: Figure translations are in progress. See original paper for figures.