Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

A Comparative study of traditional model selection methods with some of the regularization methods

Mohammed H. AL-Sharoot Amal H. Alwan
Email:mohammed.alsharoot@qu.edu.iq statistics.24.4@qu.edu.iq
University of AL-Qadisiya

Received: 28/12/2025
Accepted: 12/1/2025
Available online: 15/3/2026

Corresponding Author: Amal H. Alwan

Abstract*:* Linear regression models are used to describe and estimate the relationship between a response variable and a set of covariates. However, if some of covariates are inactive in the regression, the estimated relationship could be unstable and unpredictable. Many methods have been used over the years to identify the active covariates in the regression. In this paper, we propose Bayesian bridge-randomized expectile regression (BBRER). We compare the proposed method with the traditional model selection methods with Lasso and adaptive Lasso methods. Simulation methods show that all methods are perform comparably, however; Lasso performs the best in 80% of the simulation studies. Real data analyses using prostate cancer data also show that Lasso is the best.

Introduction: Linear regression model is used to describe the relationship between an outcome and a set of covariates. However, in the case if there are many covariates in the model, the regressio
problem could be unstable, inaccurate and unpredictable. For these reasons, different variable selection methods have been proposed of the years. See for example the traditional subset selection methods, Cp (Mallows, 2000), AIC (Akaike, 2003), AICc (Hurvich and Tsai, 1989), BIC (Schwarz, 1978) and DIC (Spiegelhalter et al., 2002). However, since the number of subsets increased exponentially with number of covariates. The traditional methods could be instable when the number of the covariates increased. To overcome this problem, regulariza- tion methods have been proposed for subset selection. See for example, Lasso (Osborne et al., 2000; Roth, 2004; Zhao and Yu, 2007) and aLasso (Friedman et al., 2010; Tibshirani, 2011; Zhao and Yu, 2006). Recently, these methods have been widely used for variable selection and estimation in linear regression mode. In this paper, we propose the BBRER in Section 2, and we compare the proposed method with the traditional model selection methods, Lasso and adaptive Lasso methods. We outline the methods in the comparison in Section 2. In Section 3, we use simulation studies to evaluate the methods, and in Section 4, we use real data to evaluate the methods. Short conclusions are given in Section 5.

1Methods

Consider the model

$$y_i = x'\beta + e_i, \quad i = 1, \bullet\bullet\bullet, n, \quad (1)$$

where yi is the outcome, xi is a set of covariates, β is a vector of coefficients and ei is the normal residual term with mean zero and variance σ2, ei ∼ N (0, σ2). used to describe and estimate the relationship between a response variable and a set of covariates. However, if some of covariates are inactive in the regression, the estimated relationship could be unstable and unpredictable. In the model (1), it is considered that some or many of covariates are assumed not active in regression. Many methods have been used over the years to identify the active covariates in the regression. See for example, the Cp criteria reported by Mallows (Mallows, 2000), however, Woodroofe (1982) and Nishii (1984) showed that this criteria tends to overfit and not consistent in selecting the true model, respectively. The Akaike information criteria (Akaike, 2003), however, Nishii (1984) showed that this criteria is not consistent. More information can be found in Vrieze (2012), Bozdogan (1987) and Yamaoka et al. (1978). The Bayesian Information Criteria (BIC, Schwarz, 1978) which is proven as a consistent, see for an example, criteria Bozdogan (1987) and Dziak et al. (2005). The deviance information criteria (DIC, Spiegelhalter et al., 2002) which is useful when the algorithms updated are available. More information on DIC can be found in Van Der Linde (2005), Zhang et al. (2019), Gao

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

et al. (2015) and Millar (2009). In this section, we propose the BBRER and review the five traditional subset selection methods (C_p, AIC, AICc BIC, DIC) in details. Also, we discuss the regularization methods (Lasso (Tibshirani, 1996) and adaptive Lasso (Zou, 2006)).

1.1 The BBRER method

Expectile regression (ER) is used to illustrate the relation between the outcome and covari- ates and has many objectives: The ER use Asymmetric Squares function (ASF) which is corresponding to a weighted OLS. When the expectile parameter (e) is equal to 0.5, the ER becoume exactly as OLS and brings the advantages of OLS. In Bayesian ER, the likelihood used given by (Newey and Powell, 1987).

$$(y\backslash\beta,\sigma^2) \propto \frac{1}{2\sigma^2}\rho e(y - x_i'\beta)^2, \qquad (2)$$

where $e \in (0, 1)$, $\rho_e = e$ if $y_i \geq \boldsymbol{x}'\boldsymbol{\beta}$ and $1 - e$ if $y < \boldsymbol{x}'\boldsymbol{\beta}$. To proceed with a Bayesian analyses, we propose the following Bridge prior distribution for $\boldsymbol{\beta}$ (Newey and Powell, 1987):

$$\rho(\beta\backslash\sigma^2) \propto exp\left\{-\lambda\left|\beta j\sqrt{\sigma^2}\right|^{\alpha}\right\} \qquad (3)$$

where λ is the regularization parameter. This prior can be written as (Mallick and Yi, 2018):

$$\frac{\lambda^{\frac{1}{\alpha}}}{2\Gamma(\frac{1}{\alpha}+1)}\exp\{-\lambda|\beta_j|^{\alpha}\} = \int_{t>|\beta_j|^{\alpha}} \frac{\lambda^{\frac{1}{\alpha}+1}}{2t^{\frac{1}{\alpha}}\Gamma(\frac{1}{\alpha}+1)} t^{\frac{1}{\alpha}}\exp^{-\lambda t}\,dt \qquad (4)$$

where, t is a mixing latent variable. We further assume has σ^2 a Gamma (a_1, a_2) prior and α has a Beta (b_1, b_2) prior. Then the hierarchical model is given by

$$t>|\beta_j|\alpha\ 2t\alpha\ \Gamma(\alpha+1)$$

$$\boldsymbol{y}|\boldsymbol{\beta},\sigma^2 \sim ASF(X\boldsymbol{\beta},\sigma^2 I_n)$$

$$\boldsymbol{\beta}|t,\sigma^2,\alpha \sim \prod_{i=1}^{k} Uniform(-\sqrt{\sigma^2}^{\frac{1}{\alpha}},\sqrt{\sigma^2}^{\frac{1}{\alpha}})$$

1

$$t \sim Gamma(\alpha+1,\lambda)$$

$$\sigma^2 \sim Gamma(a_1, a_2)$$

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

$$\alpha \sim (b_1, b_2)$$

Under the above hierarchical model, we can update the parameters from the full conditional distribution as follows:

- Updating $\boldsymbol{\beta}$

$$\boldsymbol{\beta} \sim N\left((X'EX)^{-1}X'Ey, \sigma^2(X'EX)^{-1}\right) \quad (5)$$

where E is a diagonal matrix with e.

- Updating σ^2

$$\sigma^2 \sim Gamma\left(\frac{n-1+p}{2}, \frac{1}{2}(y-X\boldsymbol{\beta})'E(y-X\boldsymbol{\beta})I\{\sigma^2 > Max_j\left(\frac{\beta_j^2}{u_j^{2/\alpha}}\right)\}\right) \quad (6)$$

- Updating u_j

$$u_j = Exp(\lambda) + \frac{|\beta_j|^\alpha}{\sqrt{\sigma^2}} \quad (7)$$

- Updating λ

$$\lambda \sim Gamma\left(a_1 + p + p/\alpha, a_2 + \sum_{j=1}^{p} |\beta_j^\alpha|\right) \quad (8)$$

- Updating α

$$\alpha \sim \alpha^{b_1-1}(1-\alpha)^{b_2-1}\frac{\lambda^{\frac{1}{\alpha}+1}}{2t^{\frac{1}{\alpha}}\Gamma(\frac{1}{\alpha}+1)} \quad (9)$$

1.2 Mallows criteria (C_p)

C_p is used to evaluate the fitting model that has been estimated using OLS and the aim is to find the correct model. A low value of C_p gives that the model is good. If p covariates are chosen from k covariates, ($k > p$), the C_p criteria is defined as:

$$C_p = \frac{SSE_p}{S^2} - N + 2(p+1), \quad (10)$$

where n is the number of the observations,$SSE_p = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_{pi}\right)^2$

package olsrr. This criteria balances between bias and variance, very simply to assess and interpret, and useful method for comparing all possible subsets of covariates.

1.3 Akaike information criteria (AIC)

AIC is a useful measure for model selection in linear regression models and other models. AIC offers a measure to compare different subsets of covariates and choose a subset that balances between goodness-of-fit and avoiding overfitting.
A low value of AIC refers to that model is good. If p covariates are chosen from k covariates, ($k > p$), the AIC criteria is defined as:

$$AIC = 2p - 2\ln(L), \tag{11}$$

where L is the maximized likelihood of the model AIC can be calculated using the R func- tion ols aic in the R package olsrr. This criteria balances between bias and variance, very simply to assess and interpret, compares non-nested subsets, simultaneous subset calculation, and useful method for comparing all possible subsets of covariates.

1.4 Corrected AIC (AICc)

Althogh AIC has a good performance in model selection, AIC tends to overfit when n is small. Thus, Hurvich and Tsai (1989) suggested a low-sample correction, leading to the following statistic:

$$AICc = AIC + \frac{2p(p+1)}{n-p-1}. \tag{12}$$

AICc can be calculated utilizing the R function AICc in the R package MuMIn.

1.5 Bayesian information criteria (BIC)

This criteria is a useful consistent measure for model selection in linear regression models and was proposed by Schwarz (1978) as a large-sample approximation to the Bayes factor. BIC

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

criteria is:

$$BIC = p \ln(n) - 2 \ln(L), \tag{13}$$

Similar to the above criterion, subset with small BIC value is selected as the preferred subset from a set of subsets. BIC can be calculated using the R function BIC in the R package stats.

1.6 Deviance information criteria (DIC)

DIC is an extension to the previous methods (AIC and BIC). It is used for subset selection in Bayesian methods when the posterior distribution is available.
This measure is defined as

$$D(\boldsymbol{\beta}) = -2 \log(p(\boldsymbol{y}|\boldsymbol{\beta})) + C, \tag{14}$$

where $\boldsymbol{y} = (y_1, ..., y_n)$ and C is a constant, $(p(\boldsymbol{y}|\boldsymbol{\beta})$ is a likelihood function. Similar to the the previous methods (AIC and BIC), model with low value 'of DIC is selected as the promise model from the 2^p models. In this section, we use DIC with the Bayesian expectile regression (BER).

1.7 Lasso

Lasso (Tibshirani, 1996) has been considerably utilize for predictor selection in linear regres- sion modes. Lasso can be defined as:

$$\text{Lasso} = \min_{\boldsymbol{\beta}} (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{15}$$

where $\lambda \geq 0$ is a regularization parameter. This criteria has been proved as a good criteria for subset selection in the regression. However, when the number of the unimportant covariates is larger than the sample size n, Lasso could be inconsistent. Lasso can be calculated utilizing the R package glmnet.

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

1.8 Adaptive Lasso (aLasso)

Adaptive Lasso (aLasso, Zou, 2006) has attracted considerable attention in the literature, es- pecially in high dimensional data. It is used for predictor selection and estimation in regression modes. aLasso can be defined as:

$$\text{aLasso} = \min_{\boldsymbol{\beta}}(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta}) + \sum_{j=1}^{p} \lambda_j |\beta_j|, \qquad (16)$$

where $\lambda_j \geq 0$ (for $j = 1, \cdots, n$) is a regularization parameters. While Lasso inconsistent estimator, Lasso is aconsistent estimator as proven by (Zou, 2006). Adaptive Lasso can be calculated utilizing the R package glmnet.

2 Simulation Studies

In this section, we used several simulated studies to evaluate the performance of the five methods (C_p, AIC, AICc, BIC, DIC) in terms of subset selection. We also compared the results with regularization approaches (BBRER, Lasso and aLasso). We listed the mean false positive numbers (MFPN, the mean of incorrectly chosen unimportant covariates) and mean false negative numbers (MFNN, the mean of incorrectly eliminated important covariates). We consider the following five simulation studies:

Simulation 1

We simulate data from the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n, \qquad (17)$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{12})$ and the covariates $\boldsymbol{x} = (x_1, \cdots, x_{12})$ are generated from $N_p(0, \Sigma)$, where Σ is the variance covariance matrix with elements $0.7^{|i-j|}$ for ith row and jth columan. We set $e_i \sim N(0, 1)$ for $i = 1, \cdots, n$ and $\boldsymbol{\beta} = (1, 3, 1, 3, 1, 3, 0, 0, 0, 0, 0, 0)$. The results are listed in Table (1). The results show that the Lasso methods performs better than the other six methods. It tends to have the smallest MFPN and MFNN. We also see that adaptive Lasso and DIC have a good performance over C_p , AIC, AICc and BIC.

Table 1: Results for Simulation 1. All results are averaged over 100 replication.

	Methods	**MFPN**	**MFNN**
Simulation 1	C_p	0.21 (0.21)	0.23 (0.31)
	AIC	0.26 (0.25)	0.19 (0.28)
	AICc	0.22 (0.31)	0.18 (0.24)
	BIC	0.18 (0.22)	0.21 (0.27)
	DIC	0.17 (0.24)	0.19 (0.24)
	Lasso	0.11 (0.17)	0.09 (0.22)
	aLasso	0.13 (0.19)	0.13 (0.23)
	BBRER	0.14 (0.19)	0.16 (0.15)

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

Simulation 2
We simulate data from the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n, \tag{18}$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{12})$ and the covariates $\boldsymbol{x} = (x_1, \cdots, x_{12})$ are generated from $N_p(0, \Sigma)$, where Σ is the variance covariance matrix with elements $0.9^{|i-j|}$ for ith row and jth columan. We set $e_i \sim N(0, 3)$ for $i = 1, \cdots, n$ $\boldsymbol{\beta} = (1, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The results are summa- rized in Table (2). The results show that the Lasso method performs better than the other six methods. The results show that Lasso is the best in term of MFPN and MFNN.

Table 2: Results for Simulation 2. All results are averaged over 100 replication.

	Methods	**MFPN**	**MFNN**
Simulation 2	C_p	0.37 (0.24)	0.11 (0.23)
	AIC	0.31 (0.27)	0.12 (0.26)
	AICc	0.34 (0.29)	0.17 (0.34)
	BIC	0.24 (0.28)	0.11 (0.32)
	DIC	0.29 (0.31)	0.10 (0.32)
	Lasso	0.19 (0.20)	0.07 (0.20)
	aLasso	0.20 (0.21)	0.09 (0.21)
	BBRER	0.21 (0.23)	0.17 (0.21)

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 014556

Simulation 3
We simulate data from the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n, \tag{19}$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{12})$ and the covariates $\boldsymbol{x} = (x_1, \cdots, x_{12})$ are generated from $N_p(0, \Sigma)$, where Σ is the variance covariance matrix with elements $0.9^{|i-j|}$ for ith row and jth columan. We set $e_i \sim N(0, 3)$ for $i = 1, \cdots, n$ $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The results are summa- rized in Table (3). The results show that Lasso is the best in terms of MFPN and MFNN.

Table 3: Results for Simulation 3. All results are averaged over 100 replication.

	Methods	**MFPN**	**MFNN**
Simulation 3	C_p	0.21 (0.31)	0.15 (0.24)
	AIC	0.19 (0.30)	0.17 (0.25)
	AICc	0.20 (0.31)	0.16 (0.21)
	BIC	0.18 (0.27)	0.14 (0.19)
	DIC	0.17 (0.29)	0.10 (0.17)
	Lasso	0.09 (0.19)	0.05 (0.16)
	aLasso	0.12 (0.20)	0.11 (0.18)
	BBRER	0.16 (0.19)	0.12 (0.20)

Simulation 4
We simulate data from the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n, \tag{20}$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{12})$ and the covariates $\boldsymbol{x} = (x_1, \cdots, x_{12})$ are generated from $N_p(0, \Sigma)$, where Σ is the variance covariance matrix with elements $0.8^{|i-j|}$ for ith row and jth columan. We set $e_i \sim N(0, 9)$ for $i = 1, \cdots, n$ $\boldsymbol{\beta} = (1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0)$. The results are summa- rized in Table (4). The results show that adaptive Lasso is the best in terms of MFPN and MFNN.

Table 4: Results for Simulation 4. All results are averaged over 100 replication.

	Methods	MFPN	MFNN
Simulation 4	C_p	0.19 (0.24)	0.12 (0.30)
	AIC	0.16 (0.27)	0.14 (0.29)
	AICc	0.17 (0.29)	0.11 (0.31)
	BIC	0.13 (0.28)	0.09 (0.28)
	DIC	0.15 (0.31)	0.11 (0.29)
	Lasso	0.07 (0.20)	0.08 (0.23)
	aLasso	0.05 (0.17)	0.07 (0.25)
	BBRER	0.10 (0.21)	0.11 (0.26)

Simulation 5
We simulate data from the following model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, \qquad i = 1, \cdots, n, \tag{21}$$

with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{12})$ and the covariates $\boldsymbol{x} = (x_1, \cdots, x_{12})$ are generated from $N_p(0, \Sigma)$, where Σ is the variance covariance matrix with elements $0.5^{|i-j|}$ for ith row and jth columan. We set $e_i \sim N(0, 25)$ for $i = 1, \cdots, n$ $\boldsymbol{\beta} = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$. The results are summarized in Table (5). The results show that the Lasso method performs better than the other six methods in terms of MFPN and MFNN.

Table 5: Results for Simulation 5. All results are averaged over 100 replication.

	Methods	MFPN	MFNN
Simulation 5	C_p	0.20 (0.26)	0.12 (0.30)
	AIC	0.14 (0.25)	0.14 (0.29)
	AICc	0.12 (0.22)	0.11 (0.31)
	BIC	0.18 (0.21)	0.09 (0.28)
	DIC	0.17 (0.22)	0.11 (0.29)
	Lasso	0.05 (0.11)	0.08 (0.23)
	aLasso	0.08 (0.19)	0.07 (0.25)
	BBRER	0.09 (0.20)	0.12 (0.27)

3 Real Data

Here, we have used the prostate cancer data which is available in the R package genridge (Friendly, 2024). This dataset have one dependent variable and 8 covariates. The dependent

variable is the the level of prostate-specific antigen. The covariates are: x_1 is the log(volume); x_2 is the log(weight), x_3 is the ages, x_4 is the log(benign), x_5 is the invasion, x_6 is the log(capsular), x_7 is a binary vector and x_8 is the Gleason score. We evaluate the seven methods using this data set. The results are listed in Table 6. The results show that the lasso method is the best. It produces the smallest MSE compared to the six methods.

Table 6: Comparison of the mean squared error (MSE) for the best model in the seven methods.

Method	MSE
C_p	0.1593
AIC	0.1442
AICc	0.1421
BIC	0.0962
DIC	0.1837
Lasso	0.0773
aLasso	0.0781
BBRER	0.0819

4 Conclusion

In this paper, we have compared the performance of the traditional subset selection methods with Lasso and adaptive Lasso. The simulation studies show that the regularization methods (BBRER, Lasso, aLasso) have a good performance compared to the other five methods C_p, AIC, AICc, BIC, and DIC. In Simulation 1, Lasso is the best. It has the smallest MFPN and the smallest MFNN compared to the others. In Simulation 2, Lasso is the best. It has the smallest MFPN and the smallest MFNN compared to the others. In Simulation 3, Lasso is the best. It has the smallest MFPN and the smallest MFNN compared to the others. In Simulation 4, adaptive Lasso is the best. It has the smallest MFPN and the smallest MFNN compared to the others. In Simulation 5, Lasso is the best. It has the smallest MFPN and the smallest MFNN compared to the others. The real data analyses support this conclusion. We have found that the Lasso method produces the smallest MSE.

References

Akaike, H. (2003). A new look at the statistical model identification. *IEEE transactions on automatic control 19* (6), 716–723.

Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika 52* (3), 345–370.

Dziak, J., R. Li, and L. Collins (2005). Critical review and comparison of variable selection procedures for linear regression. *State College, PA: Pennsylvania State University*.

Friedman, J., T. Hastie, and R. Tibshirani (2010). A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736* .

Friendly, M. (2024). *genridge: Generalized Ridge Trace Plots for Ridge Regression*. R package version 0.8.0.

Gao, G., W. Yao, K. Xia, and Z. Li (2015). Investigation of the rate dependence of fracture propagation in rocks using digital image correlation (dic) method. *Engineering Fracture Mechanics 138*, 146–155.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76* (2), 297–307.

Mallick, H. and N. Yi (2018). Bayesian bridge regression. *Journal of applied statistics 45* (6), 988–1008.

Mallows, C. L. (2000). Some comments on cp. *Technometrics 42* (1), 87–94.

Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes' factors. *Biometrics 65* (3), 962–969.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing.
Econometrica: Journal of the Econometric Society, 819–847.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, 758–765.

Osborne, M. R., B. Presnell, and B. A. Turlach (2000). On the lasso and its dual. *Journal of Computational and Graphical statistics 9* (2), 319–337.

Roth, V. (2004). The generalized lasso. *IEEE transactions on neural networks 15* (1), 16–28. Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology) 64* (4), 583–639.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58* (1), 267–288.
Tibshirani, R. J. (2011). *The solution path of the generalized lasso*. Stanford University. Van Der Linde, A. (2005). Dic in variable selection. *Statistica Neerlandica 59* (1), 45–56.
Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods 17* (2), 228.
Woodroofe, M. (1982). On model selection and the arc sine laws. *The Annals of Statistics*, 1182–1194.
Yamaoka, K., T. Nakagawa, and T. Uno (1978). Application of akaike's information criterion (aic) in the evaluation of linear pharmacokinetic equations. *Journal of pharmacokinetics and biopharmaceutics 6* (2), 165–175.
Zhang, X., J. Tao, C. Wang, and N.-Z. Shi (2019). Bayesian model selection methods for multilevel irt models: A comparison of five dic-based indices. *Journal of Educational Mea- surement 56* (1), 3–27.
Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research 7*, 2541–2563.
Zhao, P. and B. Yu (2007). Stagewise lasso. *The Journal of Machine Learning Research 8*, 2701–2726.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101* (476), 1418–1429.