Automatically Identifying Relevant Variables for Linear Regression with the Lasso Method: A Methodological Primer for its Application with R and a Performance Contrast Simulation with Alternative Selection Strategies

Sebastian Scherr [a] and Jing Zhou [b]

[a]School for Mass Communication Research, KU Leuven, Leuven, Belgium; [b]Research Center for Operations Research and Business Statistics (ORSTAT), KU Leuven, Leuven, Belgium

ABSTRACT

The abundance of available digital big data has created new challenges in identifying relevant variables for regression models. One statistical problem that gained relevance in the era of big data is high-dimensional statistical inference, when the number of variables greatly exceeds the number of observations. Typically, prediction errors in linear regression skyrocket when the number of included variables gets close to the number of observations, and ordinary least squares (OLS) regression no longer works in a high-dimensional scenario. Regularized estimators as a feasible solution include the Least Absolute Shrinkage and Selection Operator (Lasso), which we introduce to communication scholars here. We will include the statistical background of this technique that combines estimation and variable selection simultaneously and helps identify relevant variables for regression models in high-dimensional scenarios. We contrast the Lasso with two alternative strategies of selecting variables for regression models, namely, a theory-based "subset selection" of variables and a nonselective "all in" strategy. The simulation shows that the Lasso produces lower and more relatively stable prediction errors than the two alternative variable selection strategies, and it is therefore recommended to use, especially in high-dimensional settings typical in times of big data analysis.

Rapid changes in communication research have yielded a much larger number of available data, not only through publicly accessible APIs or web scraping techniques, but also through finer measurement techniques, including ecological momentary assessment (Scherer, Bickham, Shrier, & Rich, 2015) or digital trace data (Boase, 2016). These developments have made data collection cheaper and datasets larger and more complex, and some of the additional digital traces that come for free now capture by far more fine-grained constructs than ever before – fantastic opportunities for communication research and the evolution of communication theory. However, as outlined by Shah, Cappella, and Neuman (2015), these developments might as well require algorithmic solutions to explore the new data-verse to address, e.g., the problem of simultaneous correlations among variables in a prediction model. This issue has become even more pressing with the identification of relevant variables becoming one of the biggest challenges in big data (van Atteveldt & Peng, 2018). However, to maximize linear regression model efficiency while maintaining a sufficient explanatory model power, computational approaches should be considered.

CONTACT Sebastian Scherr sebastian.scherr@kuleuven.be School for Mass Communication Research, KU Leuven, Parkstraat 45, 3000 Leuven, Belgium.

Typically, the selection of variables for a statistical model is guided by hypotheses that have been formulated based on theoretical reasoning and/or previous empirical evidence or expert experience. However, given the abundance of digital trace data that are available and can be combined with other data sources (Boase, 2016), data analysis will increasingly include algorithmic solutions (Shah et al., 2015), one of which we would like to introduce and showcase here: the Least Absolute Shrinkage and Selection Operator, also known as the Lasso (Tibshirani, 1996, 2011). We will introduce the statistical background of the Lasso and then turn to R-packages, with which the Lasso can be easily applied. Our goal is to present all the necessary background information about this methodology here so it can be employed by those interested. Finally, we included scenarios in which the Lasso can (or should) be used optimally and showcased the application of the Lasso using simulation data, after which we will conclude with its strengths and pitfalls.

The Least Absolute Shrinkage and Selection Operator (Lasso) – A Primer

The Lasso method is especially useful when the number of available variables for a regression model is very high, or even higher than the number of observations, in other words whenever it is hard to differentiate between relevant and irrelevant variables. In such cases, the Lasso is a useful statistical method that automatically removes redundant variables without additional explanatory power in order to reduce model overfitting and obtain an interpretable model (see also Fonti & Belitser, 2017). In other words, the Lasso is a regression method that aims to maximize the sparsity of regression models in statistical terms, but also with regard to computation gains. In a simple linear regression with standardized variables and a centered response, the Lasso minimizes the regularized sum of squared errors (Tibshirani, 2011). It uses shrinkage techniques – a combination of the squared difference between the true and estimated values and a penalty that is proportional to the absolute value of the sum of all model coefficients. That is, a coefficient can become zero and therefore the corresponding variable can be eliminated from the regression model, thereby making the linear model easier to interpret. The Lasso combines estimation and variable selection simultaneously and is therefore especially useful for fitting a wide variety of models and large data problems with regard to the number of variables (p), observations (n), or both (Tibshirani, 2011).

Suppose that we have n observations $(X_i, Y_i), i = 1, \ldots, n$ generated from the linear model $Y_i = \alpha + X_i\beta + \varepsilon_i$, where $X_i = \left(X_{i,1}, \ldots, X_{i,p}\right)$ with p denotes the number of variables, the linear regression error ε_i follows a normal distribution $N(0, \sigma^2)$, β denotes the regression coefficient vector $\beta = \left(\beta_1, \ldots, \beta_p\right)^\top$, and α denotes the intercept. In a linear regression where the number of observations n is typically much larger than the number of variables p, the ordinary least squares (OLS) estimator has been widely used due to its accuracy and efficiency.[1] However, in times of big data, the number of variables might as well be larger than the number of observations (i.e., $p > n$ settings), and this is where traditional OLS is not only not well-defined but also no longer applicable. For instance, microarray data usually contain p = 10K gene expressions that may or may not be associated with a disease captured for only n = 100 patients or p different biopsychosocial indicators of depression that are combined with clinical indicators all aiming to explain depression among only n = 50 patients. To perform the estimation and select relevant variables at the same time, Tibshirani (1996) developed the Lasso method for such settings, as it minimizes the l_1 regularized least-squares loss and it is defined as:

$$\left(\hat{\alpha}, \hat{\beta}\right) = argmin_{(\alpha,\beta)} \left\{ \sum_{i=1}^{n} \left(Y_i - \alpha - \sum_j \beta_j X_{ij} \right)^2 + \lambda \sum_j |\beta| \right\}$$

where the shrinkage parameter λ is chosen by the user. On the one hand, a large λ shrinks most coefficients to zero, while the intercept α remains unchanged and leaves a sparse model solution that makes the regression model more interpretable. On the other, a small λ introduces less bias to the estimator, and setting

$\lambda = 0$ leads to an unbiased OLS estimator. In other words, the Lasso filters for similar variables, but when $\lambda = 0$, the equation is equal to an OLS estimation, and therefore the Lasso can be used within a feature selection task (see Fonti & Belitser, 2017 as a further reading) – the identification of the most influential variables in a regression model. The Lasso returns a sequence of different models for different values of λ, which inform the researcher about the most suitable variables in a regression; this process can be automatized within a machine learning task. With a proper choice of λ, the Lasso minimizes the prediction error of a regression model while recovering the underlying sparse regression model.

Of note, the Lasso selects at most n out of p predictor variables before it saturates (Efron, Hastie, Johnstone, & Tibshirani, 2004) in the $n > p$ case due to the nature of convex optimization. This "select at most n variables" feature naturally becomes a limitation of the Lasso for a variable selection method. Additionally, if there is a group of variables with high pairwise correlations, the Lasso tends to randomly select one variable from the group. The abovementioned limitation makes the Lasso less suitable for selecting highly correlated variables, since the ideal variable selection method should be able to include the whole group of highly correlated variables once one variable among those is selected. An example of highly correlated variables is binary (i.e., 0, 1) dummy variables coded from categorical variables with several levels. In such cases, alternative regularization techniques such as *elastic net* (Zou & Hastie, 2005) or *group Lasso* (Yuan and Lin (2007) should be considered. Details about the alternatives can be found in the discussion section.

Since its introduction, the Lasso (Tibshirani, 1996) has become increasingly relevant for high-dimensional statistics (i.e., when the number p of variables greatly exceeds the number n of observations), which is why it is of particular relevance in times of big data and computational (social) sciences. The Lasso is a machine-learning procedure that combines model selection and model shrinkage, is optimized for out-of-sample prediction rather than within-sample explanation, and has been advocated as an alternative to OLS regression in theory building and hypothesis testing (Hindman, 2015). The biggest advantage might be that machine learning-based techniques, such as the Lasso, estimate more parsimonious statistical models, thereby alleviating problems of overfitting that are even more likely to occur in very large datasets. We therefore focus on the Lasso as a data analytical method to improve theory testing in the social sciences in times of big data and illustrate the advantages of this newer, machine learning-based method over traditional linear regression analysis using simulated data.

However, several cautions need to be mentioned since variable selection techniques are not only prone to selection errors and sensitive to data changes, but also do they not automatically offer p-values for the selected variables. Additionally, variables without explanatory power can be accidentally included or relevant variables can be excluded. These two types of selection errors may either lead to model *overfit* or *loss of model identification power*. Importantly, the Lasso will accurately identify strong variables (i.e., it will address the problem of accidental variable exclusion). However, there might as well be small non-zero variables that may be of importance within a model, which might be overseen by the Lasso. Keeping these abovementioned challenges in mind it is important that researchers are aware of this uncertainty when constructing a model. One approach beyond various statistical approaches (see Belloni, Hansen, & Chernozhukov, 2014) would be to rely on strong, theory-driven reasoning that prevents researchers from accidentally overlooking or excluding relevant variables also in times of changing big data.

Hence, adjusting a model's complexity by in- or excluding available variables in order to achieve a good overall model fit can be addressed by the researcher (e.g., by taking simply all or only certain variables out of all data available especially when gathered as part of a larger study). However, automatic and thus more reproducible strategies to adjust the number of free parameters in models including machine learning and/or regularizers have also been suggested. All of this taps into Occam's razor, a philosophical principle that advocates the more parsimonious explanation for the same occurrence (if there are alternative ones; e.g., Rasmussen and Ghahramani (2001)). Importantly, both a too high and a too low model complexity reduce test performances (Rasmussen & Ghahramani, 2001). Reversely, if a more sophisticated explanation does a better job, then it should be given priority over the simpler one. Hence, we will consider three scenarios in this paper, and will

compare the Lasso with two alternative strategies of selecting variables for linear regression as part of a performance contrast simulation.

Method

The performance contrast simulation has been conducted using R. To approximate the numerical solutions of the Lasso, we used the built-in R function *cv.glmnet*, which tunes the shrinkage parameter λ using K-fold cross-validation, in the R package glmnet (Friedman, Hastie, & Tibshirani, 2009).[2] The glmnet package, allowing both linear and generalized linear models, implements the elastic-net regularizer (Zou & Hastie, 2005), which is a mixture of the l_1 (corresponds to the Lasso regularizer) and l_2 regularizers (the Ridge regularizer).[3]

The first scenario ("all in") reflects a situation in which a researcher deciphers a social phenomenon and breaks it up into different subdimensions, decides ex ante which variables are meaningful indicators for all of these subdimensions, collects them, and then simply includes all of them in a (most likely overfitted) regression model. In order to simulate this scenario, we included between 7 and 40 variables. The second scenario ("subset selection") speaks to the situation in which a researcher has collected variables that are indicative of different subdimensions of the same social phenomenon, e.g., as part of a larger study, but, for theoretical reasons, decides to include only some of them as meaningful indicators to the model (i.e., partial variable inclusion). To simulate it, we always included 7 variables, while the total number of variables ranged from 7 to 100. The seven variables in our models were composed as follows: the first two variables had a non-zero true regression coefficient, and the remaining five were randomly chosen. Hence, the first two variables represent two meaningful indicators since they are of importance to the true regression model and the remaining five variables in the model represent the random aspect of the "subset selection" (i.e., a researcher might not have all the information needed in order to decidedly select the remaining five variables). In order to control variance resulted from variable selection uncertainty, the subset ofvariables has only been selected once using seed number 2 (a pseudo-random number that is often used to produce the same numerical results; i.e., a fixed decision on the subset of the variables over simulation repetition for a fixed number of variables).

The third scenario is the described Lasso method. The central outcome used to compare and contrast the performance of the different scenarios was the median prediction mean squared error (MSE) presented below in Figure 1 in relation to the total number of variables included in the model. The number of variables in the regression models was set to vary between 7 and 100. The simulation was repeated 1,000 times by generating 1000 new datasets for a fixed number of variables. For each simulation repetition, the three variable selection strategies are performed, and the corresponding estimations are recorded for further comparison. Therefore, the performance differences between the different variable selection strategies become clear.

More specifically, in all scenarios we assumed the first $p-2$ variables to be independent and following the standard normal distribution and the last two variables to follow a Bernoulli distribution with success probabilities of 0.5 (e.g., gender) and 0.4 (e.g., an income measure with a specific threshold that only 40% of the sample reach), respectively. Variables following a Bernoulli distribution have the values 0, 1 with 1 being the "success probability". We follow the simulation protocols from Zou and Hastie (2005; example a) and Tibshirani (1996; example 1), in which the true regression coefficient vector is taken to be $\beta = \left(\beta_1, \ldots, \beta_p\right)^\top = (3, 1.5, 0, 0, 2, \ldots, 2, 1)^\top$, with all entries ranging from 6 to $p-2$, being zero; the intercept α is neglected. In each simulation repetition, the values of the variables and the true regression coefficient are randomly generated using a fixed seed number and they remain unchanged, and the regression error ε is generated under a different seed number with standard deviation $\sigma = 1$. In addition, for each simulation repetition, we followed Zou and Hastie (2005) and generated 240 samples, of which 40 were used to estimate the regression coefficient vector and the remaining 200 samples were used for calculating the out-of-sample prediction error, defined as:

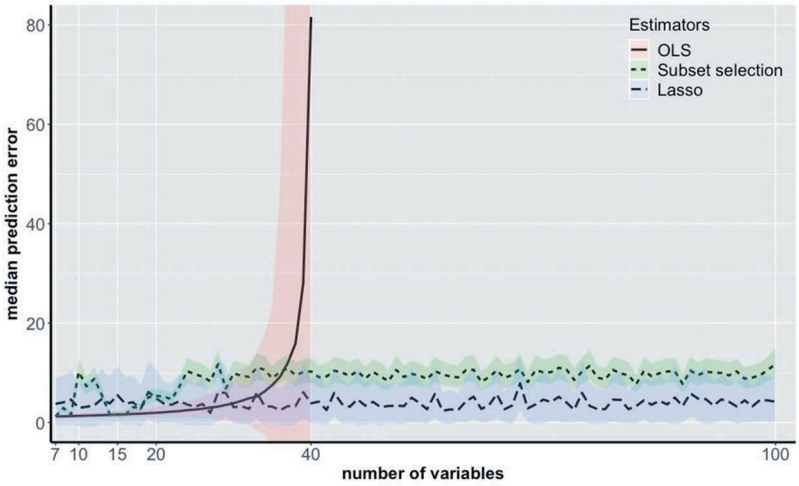

Figure 1. Median out-of-sample prediction error of three different variable selection strategies ("all-in" OLS regression; "subset selection", and Lasso method). The figure shows the simulation for a fixed sample size of $n = 40$ and the number of variables in the regression models ranging from 7 to 100 following the examples of Zou and Hastie (2005; example a) and Tibshirani (1997; example 1) with a slight modification: two dichotomous variables have been incorporated as well. The first $p - 2$ variables are independent and follow a standard normal distribution; the last two variables follow a Bernoulli distribution with success probabilities of 0.5 (e.g., gender) and 0.4 (e.g., an income measure with a specific threshold that only 40% of the sample reach), respectively. Variables following a Bernoulli distribution were 0, 1 valued with the probability of taking value 1 being the "success probability." Prediction errors become especially high in OLS regression, when the number of variables included in the regression model gets close to the number of observations (here $n = 40$; see line). The simulation of the two alternative variable selection strategies (two dotted lines) shows that the Lasso method produces relatively stable prediction errors that are typically lower than in the "subset selection" scenario, especially if the number of variables is $p > 15$.

$$PE\left(\hat{\beta}\right) = \sum_{i=1}^{n_{test}} \left(Y_{test,i} - X_{test,i}\hat{\beta}\right)/n_{test}$$

Because the number of samples used for the error estimation was set to 40, only the second ("subset selection") and third (Lasso) scenarios are reflective of high-dimensional statistics (i.e., when the number of variables exceeds the number of observations, i.e., $p > n$).

Results

The simulation shows that in cases where $n \geq p$, the OLS estimator mostly has the lowest median prediction MSE among the three different estimator scenarios – except for cases in which the number of variables is close to the number of observations. However, in cases where $p > n$ and the OLS estimator is no longer applicable, the Lasso clearly outperforms the "subset selection" with regard to the median prediction MSE. The standard deviation of the prediction MSE of 1,000 repetitions of the Lasso is higher than its counterparts, as the Lasso might select different variables in the 1,000 repetitions and the standard deviation reflects this potentially higher variability in the selected variables (see Table 1).

Importantly, the median prediction MSE of the Lasso is comparatively stable, though the number of variables varies largely (see also Figure 1). Table 1 shows the estimated error results for the estimators in each of the three scenarios.

Table 1. Median prediction errors (and standard deviations) of 1,000 simulation repetitions of regression models with different strategies to identify relevant model variables (p).

	$p = 7$	$p = 8$	$p = 9$	$p = 10$	$p = 15$	$p = 20$	$p = 30$	$p = 100$
Scenario 1:	1.186	1.231	1.266	1.310	1.577	1.933	3.862	–
"All in" OLS regression	(0.184)	(0.197)	(0.217)	(0.243)	(0.364)	(0.545)	(1.982)	
Scenario 2:	1.186	2.902	1.575	10.200	1.570	5.339	9.865	11.717
"Subset selection"	(0.184)	(0.397)	(0.247)	(1.264)	(0.241)	(0.683)	(1.185)	(1.572)
Scenario 3:	3.731	4.030	4.584	2.910	5.471	4.838	3.054	4.162
Lasso Method	(2.451)	(2.529)	(2.757)	(2.188)	(2.868)	(2.793)	(2.025)	(2.195)

Numbers are the median out-of-sample prediction errors based on 1,000 simulation repetitions with the corresponding standard deviations shown in brackets.

The simulation data are generated from the linear model following the examples of Zou and Hastie (2005; example a) and Tibshirani (1997; example 1) with a slight modification, where the first $p - 2$ variables are independent and standard normal distributed; the last two variables follow the Bernoulli distribution with a success probability of 0.5 and 0.4, the true regression coefficient vector is taken to be $\beta = (\beta_1, \ldots, \beta_p)^\top = (3, 1.5, 0, 0, 2, \ldots, 2, 1)^\top$ with all entries ranging from 6 to $p - 2$ being zero, and the intercept α is neglected.

To observe the impact of the number of variables p on the performance of the three scenarios, we varied p from 7 to 100 and randomly generated 1,000 datasets for a fixed p. In each simulation repetition under a fixed p, we generated 240 samples, of which 40 were used to estimate the regression coefficient vector and the remaining 200 samples were used for calculating the out-of-sample prediction error. To obtain the Lasso estimations, we used the *R* package *glmnet*.

The exact replication code is available under: https://osf.io/7fqjr/

Discussion

The rapidly changing field of communication research is facing the situation that not only is a much larger amount of data now available through digital research methods, but these data are also moreover combinable with traditional data; thus, research opportunities have grown exponentially (e.g., Boase, 2016). Digital research methods and the vast amount of data offer completely new ways of looking at new but also established communication phenomena – great times for communication research. However, Shah et al. (2015) recently pointed to the necessity that digital developments might as well require algorithmic solutions to handling data modeling. One fundamental issue in times of computational social science and big data is the abundance of variables, while simultaneously maximizing model efficiency and keeping the explanatory power of the model sufficiently high (van Atteveldt & Peng, 2018). Traditionally, the selection of variables is hypothesis-driven, but algorithmic alternatives could support researchers in conjunction with traditional reasoning in a changing research environment (Shah et al., 2015).

In this paper, we addressed this pressing issue and introduced a computational solution for it – the Lasso (Tibshirani, 1996, 2011). After a brief introduction of the method, which can be implemented using R, we presented the results of a performance contrast simulation, in which we compared the Lasso method with two alternative scenarios of how variables are usually selected for linear regression: a theory-driven selection of meaningful variables (i.e., "subset selection") and the inclusion of all available variables (i.e., "all in").

The simulation showed that an OLS estimator typically has the lowest median prediction MSE if the number of cases (n) exceeds the number of variables (p). However, in the case of $p > n$, which is of increasing relevance in times of big data, not only is the OLS estimator is no longer applicable, but the Lasso estimator also clearly outperforms alternative variable selection strategies.

Strengths and Limitations

As with all studies, some important aspects must be kept in mind when interpreting the findings. First, it is important to keep in mind that the regularizer of the Lasso itself causes bias; for the classical linear regression setting where $n > p$ with fixed p, the Lasso estimator should not become the default choice and it should always be used with caution. Second, the Lasso can only select one variable in a group of correlated variables; hence, if, for example, a multi-categorical variable has been split into several binary indicators, the Lasso would select only one level. Finally, the Lasso selects at most n variables upon

saturation. However, Yuan and Lin (2007) proposed the *group Lasso*, which uses an additional group regularizer for group selection, while Zou and Hastie (2005) suggested an elastic net regularizer (i.e., a mixture of the l_1 and l_2 penalties) that allows selection between correlated variables without restricting the maximum number of selected variables to be equal to the sample size. More recently, Gertheiss and Tutz (2010) further inspected the selection of categorical variables; especially, the order of the levels of the ordinal variables is considered.

For researchers who want to use the Lasso it is important to point out situations in which the Lasso method could be misused. First, the Lasso was originally developed for the linear model assuming a linear relationships between predictor variables and the response variable, but later was extended to other models (e.g., generalized linear model, cox model, etc.) assuming nonlinear relationships. Researchers can therefore adjust the Lasso accordingly (see https://osf.io/7fqjr/). And second, the Lasso could promote questionable research practices including "data peeking", dropping a response variable, dropping an experimental condition, and controlling for (only certain) predictor variables (Simonsohn, Nelson, & Simmons, 2014). In order to avoid overfitting of the data, researchers might use the Lasso simply to trade-off between variance and bias without considering underlying theoretical mechanisms. A theory primacy is therefore strongly recommended at all times.

Conclusion

Digital methods in (computational) social sciences will more often result in large and detailed datasets in which the number of variables exceeds the number of cases. In such cases where a traditional OLS estimator can no longer be used, regularized estimators should be used. Relatedly, an increasingly problematic aspect within this rapidly changing field is the selection of meaningful variables for regression-based prediction models. To overcome this specific obstacle in such paradigm-shifting times for social science, we introduce the Lasso method and show that a Lasso-based selection of variables for a regression model outperforms other strategies of selecting relevant variables with regard to the median prediction error. Researchers should consider the Lasso method an alternative strategy to reduce datasets and identify meaningful variables when the number of variables exceeds the number of observations.

Notes

1. In an OLS regression, the estimator is unbiased and reaches the lower bound of the variance with $\hat{\beta} = (X^\top X)^{-1} X^\top Y$, which comes from the unique minimizer of the quadratic function $\sum_{i=1}^{n} \left(Y_i - \sum_j \beta_j X_{ij}\right)^2$. Of note, the first column of the design matrix corresponds to the regression intercept.

2. K-folds cross-validation randomly splits the original dataset into equal-sized K subsets. Each time, it uses $K-1$ subsets for the estimation before the validation is performed using the last subset to predict the out-of-sample prediction MSE. Because there are K possible validation sets, for the final prediction, the MSE is the average over K times validation (i.e., $\sum_{k=1}^{K} PE(\beta_k)/K$).

3. The Lasso estimator is determined by setting the mixing parameter of the mixed regularizer to 1.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Sebastian Scherr http://orcid.org/0000-0003-4730-1575
Jing Zhou http://orcid.org/0000-0002-8894-9100

References

Belloni, A., Hansen, C., & Chernozhukov, V. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 1–23. doi:10.1257/jep.28.2.1

Boase, J. (2016). Augmenting survey and experimental designs with digital trace data. *Communication Methods and Measures*, *10*(2–3), 165–166. doi:10.1080/19312458.2016.1150975

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regresssion. *The Annals of Statistics*, *32*(2), 407–499. doi:10.1214/009053604000000067

Fonti, V., & Belitser, E. (2017). *Feature selection using LASSO*. Research Paper in Business Analytics. VU Amsterdam. Retrieved from https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf

Friedman, J., Hastie, T., & Tibshirani, R., (Producer). (2009). *Glmnet: Lasso and elastic-net regularized generalized linear models*. [R package]. Retrieved from http://CRAN.R-project.org/package=glmnet

Gertheiss, J., & Tutz, G. (2010). Sparse modeling of categorial explanatory variables. *The Annals of Applied Statistics*, *4* (4), 2150–2180. doi:10.1214/10-AOAS355

Hindman, M. (2015). Building better models: Prediction, replication, and machine learning in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 48–62. doi:10.1177/0002716215570279

Rasmussen, C. E., & Ghahramani, Z. (2001). Occam's razor. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13, pp. 294–300). Cambridge, MA: MIT press.

Scherer, E. A., Bickham, D. S., Shrier, L. A., & Rich, M. (2015). Evaluating multiple intensively collected media use measures: Validity and reliability of momentary assessments. *Communication Methods and Measures*, *9*(3), 170–187. doi:10.1080/19312458.2015.1061653

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 6–13. doi:10.1177/0002716215572084

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. doi:10.1037/a0033242

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*, *58*(1), 267–288. doi:10.1111/rssb.1996.58.issue-1

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *73*, 273–282. doi:10.1111/j.1467-9868.2011.00771.x

van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92. doi:10.1080/19312458.2018.1458084

Yuan, M., & Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, *94*(1), 19–35. doi:10.1093/biomet/asm018

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. doi:10.1111/j.1467-9868.2005.00503.x