Comments on: High-dimensional simultaneous inference with the bootstrap

Hanzhong Liu & Bin Yu

TEST
An Official Journal of the Spanish Society of Statistics and Operations Research

ISSN 1133-0686
Volume 26
Number 4

TEST (2017) 26:740-750
DOI 10.1007/s11749-017-0559-x

Springer

Your article is protected by copyright and all rights are held exclusively by Sociedad de Estadística e Investigación Operativa. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Springer

Author's personal copy

TEST (2017) 26:740–750
DOI 10.1007/s11749-017-0559-x

DISCUSSION

Comments on: High-dimensional simultaneous inference with the bootstrap

Hanzhong Liu[1] · **Bin Yu**[2,3]

Published online: 9 October 2017
© Sociedad de Estadística e Investigación Operativa 2017

Abstract We provide comments on the article "High-dimensional simultaneous inference with the bootstrap" by Ruben Dezeure, Peter Buhlmann and Cun-Hui Zhang.

Keywords Ranking and selection · Mean squared error · Coverage · High-dimensional statistical inference · Bootstrapping LassoOLS

Mathematics Subject Classification 62F07

1 Introduction

We congratulate the authors on their interesting and thought-provoking paper (called the DBZ paper thereafter) and appreciate the Editor's invitation to discuss. This paper

This comment refers to the invited paper available at: doi:10.1007/s11749-017-0554-2.

This work is supported in part by NSF Grant DMS-1228246, ONR Grant N00014-16-1-2664, AFOSR Grant FA9550-14-1-0016, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant Agreement CCF-0939370 (to Yu).

Electronic supplementary material The online version of this article (doi:10.1007/s11749-017-0559-x) contains supplementary material, which is available to authorized users.

✉ Bin Yu
binyu@berkeley.edu ; binyu@stat.berkeley.edu

1 Department of Industrial Engineering, Center for Statistical Science, Tsinghua University, Beijing, China

2 Department of Statistics, University of California, Berkeley, CA, USA

3 Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA

Author's personal copy

addresses the difficult problem of high-dimensional statistical inference in the context of sparse linear models and Lasso-related methods. Inference or uncertainty quantification is central to statistics and at the frontier of modern statistics and is increasingly attracting attention in machine learning and data science. The challenge of quantifying uncertainty is particularly relevant in big data settings. That is, it commonly arises in fields such as neuroscience where researchers are generating terabytes of fMRI data with every experiment, in genomics where an individual's genome can be sequenced and analyzed in the blink of an eye, in social science where social media data allow understanding into social influence at an unprecedented scale, and in astronomy where the arrival speed of data from telescopes makes online decisions unavoidable. In these fields, data-driven scientific discoveries are becoming an accepted alternative to the more traditional hypothesis-driven discoveries.

As we dive deeper and deeper into a realm in which questions are both raised and answered on the basis of data analysis, it is vital to take a moment to pause, take a step back, and ponder two key questions that highlight the validity of the answers derived in our data-driven world:

1. How is uncertainty quantification used in these new data problems? For these new data problems, it is still the case that a very uncertain answer should not be taken seriously while a precise one deserves attention and resource to follow up.
2. What are effective metrics to assess whether an inference method is adequate or good, in addition to coverage? More metrics are necessary because the effect size of a new medical treatment is not adequately reflected in the coverage. Furthermore, although ranking of the treatments is very useful for a patient, it is not addressed by the coverage either.

At this point, we would like to note that our discussion surrounding these questions will be strongly relevant for a wide variety of inference problems, reaching far beyond the sparse linear model setting we focus on here.

To address the questions above, we propose two important metrics for assessing the performance of an inferential procedure such as a confidence interval: MSE and ranking. We will present some preliminary comparisons (in terms of MSE, coverage and ranking) of various methods covered in the DBZ paper based on two data-inspired sparse linear model simulations. Finally, in writing this discussion, we aim to stimulate further discussions with the authors and others.

2 Statistical inference for informed decision making

In our view, one of the most important (if not the most important) contributions of statistical inference is to help make informed decisions with appropriately quantified uncertainties. To ground our discussion, we would like to describe a long-term interdisciplinary project with the Gallant Neuroscience Lab at UC Berkeley that studies primate visual pathways. For a particular voxel in a human subject's visual cortex V1, fMRI brain signals (indirect correlates of aggregated neuron activities) were measured in the Gallant Lab. A predictive model has been built via the Lasso based on features of images after Gabor transformations that are known to mimic V1 neurons.

Author's personal copy

One scientific question to ask is: which Gabor features are driving the voxel response among 10,921 features? A good statistical data analysis should help narrow down (or recommend) possible driving or important features (and ultimately experiments have to be run to validate such recommendations in the Gallant Lab).

2.1 Effect size and statistical confidence

The ability to recommend important predictors is a key feature that has led to the success of data analytics in making data-driven discoveries in science and business alike. Many of these analyses are based on linear models which allocate an estimated coefficient to each feature. It is both natural and informative to look at the size of the estimated coefficients for the features as a measure of "effect size" even before a measure of uncertainty (as scientists and statistics practitioners always do). The effect size can be used to assess domain importance or domain significance, while the effect size normalized relative to a uncertainty measure provides statistical significance. While interpretations of both the raw coefficients themselves and the normalized versions can differ, they are both important for down-stream decision making.

In the fMRI project, the effect size of a Gabor feature as measured by the estimated coefficient multiplied by the size of that Gabor feature indicates the contribution of that Gabor feature to the voxel response. For Gabor features with small average effect sizes, the small effect sizes can be too small to be of interest, no matter how large the normalized effect sizes are (or how statistically significant the results are). That is, not only do we need to seek to surpass a statistical significance threshold for the normalized coefficient, we must also target a domain significance threshold for the raw coefficient.

The more traditional threshold of statistical significance is a direct assessment of importance relative to the uncertainty due to noise in the process of data generation. While the statistical significance threshold is notoriously domain-agnostic (in that a threshold of 0.05 is universal), a domain-knowledge-driven threshold for the effect size would be strongly domain dependent. For example, in some applications a raw coefficient of 2.5 is considered large, whereas in others a coefficient of 500 is considered large. Moreover, in many settings, we do not always know what is considered large, for example because the technology used to generate the data (e.g., fMRI) is relatively new.

To evaluate an inference method by considering whether an unnormalized effect size passes a domain significance threshold, as described above, we propose the MSE as a suitable metric. The MSE estimates how close an estimator (the estimated effect size) is to the parameter value (true effect size), incorporating both bias and variance. More precisely, the MSE is directly concerned with this estimated effect size, or raw unnormalized coefficient value. We believe that, in addition to coverage, it is vital to include MSE as a metric for inference method evaluation. Moreover, MSE directly evaluates effect size estimation than the size of a confidence interval. The latter is closely related to the standard deviation (or variance) of an estimator. Like MSE, coverage deals with bias and variance together, but in an implicit way (unlike in MSE).

Author's personal copy

2.2 Ranking important features or covariates

Among the features that are recommended by a statistical investigation, being able to arrange the features in order of importance is crucial. For example, in a gene–disease association study, limited resources require researchers to pre-specify a number of genes for further study; therefore, it seems more compelling to discover the most k important genes than to test the significance of each gene. In this situation, ranking the coefficients of each of the covariates is much more important than covering the coefficients in high-dimensional models. In other words, from the decision-making point of view, a more basic problem is to identify, for example, the best k or the worst k among the features or covariates under consideration with a high probability of a correct selection.

Ordering statistical findings to support decision making can be traced back at least to the 1950s under the name of selection and ranking, including fundamental and excellent work (Bahadur 1950; Bahadur and Robbins 1950; Bechhofer 1954; Bechhofer et al. 1968). As a leading researcher of selection and ranking, Professor Milton Sobel pointed out (Mukhopadhyay 2000) "this new area would be a revolution in the sense of replacing the general overuse of 'Testing the Hypothesis of Equality' by new decision-theoretic models for ordering populations with prescribed confidence in the resulting decision".[1] This work studied ranking and selecting the top k populations with the highest population means from a total of p populations, where each population was observed with independent observations.

In our modern setting of high-dimensional inference, we often want to rank features that are dependent. For example, the Gabor wavelet features from our fMRI project do not form a set of independent features, implying that the existing selection and ranking methods do not directly apply. Work needs to be developed to encompass these modern problems in our selection and ordering setting.

2.3 Inference methods via bootstrap for sparse linear models

Let us now turn to the setting of DBZ and the world of high-dimensional sparse linear models. In this section, we will review various inference methods based on the bootstrap. In the DBZ paper, uncertainty is assessed on the estimated parameters in a sparse linear model under three bootstrap methods: bootstrapping the de-biased Lasso procedures including a residual bootstrap, a multiplier wild bootstrap and a special version of a paired bootstrap to construct individual confidence interval and simultaneous confidence regions for parameters in high-dimensional linear models. The de-biased Lasso is designed to remove the bias of the Lasso and to derive an asymptotically normal distributed estimator for each individual coefficient in high-dimensional sparse linear models with Gaussian and homoscedastic errors. Bootstrapping the de-biased Lasso can deal with non-Gaussian and heteroscedastic errors. Both empirical results and theoretical investigations show the

[1] In 1950s, many statisticians including Jack Kiefer, Herbert Robbins, Jack Wolfowitz, R. C. Bose, Abraham Wald and Milton Sobel thought that the area of ranking and selection is a "revolution".

Author's personal copy

advantage of the proposed methods. In particular, these methods perfectly control the type I error and the familywise error rate at the expense of losing some power and avoid the "beta-min" condition that seems to be needed by bootstrapping the adaptive Lasso and bootstrapping the LassoOLS as in our earlier paper (Liu and Yu 2013), and the super-efficiency phenomenon will not occur. However, the confidence intervals constructed by the new proposals of DBZ sometimes are large for small coefficients.

Bootstrapping the Lasso (Chatterjee and Lahiri 2010, 2011)[2] or LassoOLS (Liu and Yu 2013) is a simpler tool, both conceptually and computationally, for high-dimensional inference in sparse linear models than the de-biased Lasso. They perform comparably in terms of coverage probability and interval length as shown in the comprehensive empirical studies in Dezeure et al. (2014). Compared with bootstrapping the de-biased Lasso, there are two advantages of bootstrapping the Lasso or LassoOLS. First, they are built on top of canonical statistical techniques, the bootstrap, the Lasso (and the OLS), which are all well known to a broad range of scientists and data scientists and hence easily accessible to them; second, they can be computational faster when the number of replications in the bootstrap B is smaller than the number of covariates p.

However, a "beta-min" condition was imposed in Chatterjee and Lahiri (2011) and Liu and Yu (2013), and bootstrapping the Lasso or LassoOLS produces confidence intervals close to zero lengths and zero coverage probabilities for small but nonzero coefficients, e.g., confidence intervals [0, 0] in extreme cases. Small coefficients or parameters (relative to noise size) are difficult to estimate, and we are curious to see data problems in the real world where such coefficients are the focus (since we have not encountered such a problem ourselves). Without such real problems, it is hard to justify why we should compare methods based on their performances on small coefficients and make recommendations to practitioners based on such comparisons. Assume such data problems exist (which could be a big assumption), we then face three choices: bootstrapping the de-biased Lasso or Lasso (or LassoOLS). None is perfect or dominates the other two. For simplicity, let's look at their behaviors for zero coefficients. Bootstrapping the de-biased Lasso gives large confidence intervals and potentially also has large biases, while bootstrapping the Lasso or LassoOLS gives small confidence intervals (sometimes zero confidence intervals) and has small bias and MSE. The coverage of the former is better because of the large intervals, but its effect size estimate is disappointing. The latter is the opposite. It seems reasonable to believe that to most people, a small confidence interval near zero means the coefficient is small and we are sure about it, while a large confidence interval near zero says that the coefficient is small, but we are not sure about it. The former seems to us a more informative evidence to use in practice.

[2] In paper (2010), Chatterjee and Lahiri showed that bootstrapping the Lasso is inconsistent whenever one or more regression coefficients are zero. They proposed bootstrapping the thresholded Lasso (Chatterjee and Lahiri 2011) and bootstrapping the adaptive Lasso (Chatterjee and Lahiri 2013) and showed their validation under the beta-min condition and other appropriate conditions. However, bootstrapping the Lasso performs comparably (but is simpler) to the other two methods in simulations where the signal-to-noise ratio is not high.

Author's personal copy

3 MSE, coverage, and ranking in data-inspired simulation studies

It is relatively straightforward to investigate MSE and coverage using simulation studies in a variety of inference settings. In what follows, we will display the MSE and coverage for each individual coefficient.

Due to correlations between covariates and subsequent dependencies between coefficient estimates, ranking and selecting covariates/coefficients in modern linear regression models are more complicated than in the traditional settings of Milton Sobel and others. Traditionally, the t-statistic or p value is used to rank the importance of covariates in a linear regression model. In the bootstrapping setup, p value cannot be estimated accurately using limited bootstrap replications, so we will not consider it for ranking purpose. Since both a coefficient estimate and its confidence interval length are related to the importance of covariates, a naive generalization of the t-statistic is the ratio the coefficient estimate and its corresponding confidence interval length. This new importance measure is proportional to the t-statistics if the coefficient estimate follows normal distribution. We can rank the covariates by the absolute values of the new importance measure. In our simulation, we consider confidence interval length of confidence levels 90, 95, and 99% and find that 99% confidence interval length works the best, which is comparable to the t-statistic. We call this ratio a generalized t-statistic. To deal with the special case of $0/0$, we rank the coefficients with zero estimates and zero-length confidence intervals at the end of the ranking sequence, meaning that they are not important at all.

To evaluate the ranking performance, we use the ReDiscovery(k), which is defined as the expected number of top k covariates being ranked within the top k again (Lu et al. 2009). The larger the ReDiscovery(k), the better the ranking.

In the next sections, we will focus on data-inspired simulation studies for comparing (1) the bias-variance trade-off, MSE and coverage of the Lasso, the de-biased Lasso and the LassoOLS, and (2) the ranking performance of bootstrapping these estimators.

3.1 Simulation study setups based on real data sets

Our simulations are based on two real data sets: fMRI data introduced earlier (Kay et al. 2008) and the Ames Iowa Housing data.[3] The fMRI recorded measurements of blood oxygen level-dependent activity at 1331 discretized 3D brain volumes ($2 \times 2 \times 2.5$ mm): cube-like units called voxels. There are 1750 observations and 10,921 covariates (Gabor features) for each voxel. In the simulation, we consider one of the voxels as the response and 2000 covariates having the top largest marginal correlations with the response. The Ames Iowa Housing data contained information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA, from 2006 to 2010. The response variable is the housing price, and there are 1543 observations and 80 covariates in a training data set. After removing 3 outliers and including some quadratic and two-way interaction terms, we finally form a design matrix of dimensions 1540×243. We will refer to this data set as

[3] http://www.amstat.org/publications/jse/v19n3/decock/AmesHousing.xls.

Author's personal copy

Housing data for simplicity. The covariates in the fMRI data are more or less normally distributed, while the covariates in the Housing data are not.

The true coefficients β^* in the simulations are computed by the Lasso for the two real data sets. The β^* has 46 and 71 nonzero elements in the two data sets, respectively. After β^* is generated, we simulate $Y = (y_1, \ldots, y_n)^T$ from a linear regression model $y_i = x_i^T \beta^* + \epsilon_i$ by generating independent error terms. The error terms are non-Gaussian and heteroscedastic, which are generated in two steps: (1) generating $\eta_i, i = 1, \ldots, n$ i.i.d. from uniform distribution on $[1, 3]$. η_i's are used to make the errors heteroscedastic, which are generated once and then kept fixed; (2) generating $\epsilon_i, i = 1, \ldots, n$ by the equation $\epsilon_i = \sigma \eta_i \frac{\xi_i - 1}{\sqrt{2}}$, where $\xi_i, i = 1, \ldots, n$ are i.i.d. from χ_1^2 distribution with one degree of freedom. We set the value of σ such that the signal-to-noise ratio (SNR) equals 0.5, 1, 2 and 5, respectively. SNRs considered here are much smaller and more realistic than those used in the DBZ paper and the review paper (Dezeure et al. 2014) where the smallest SNR is around 3.5 and the largest one is more than 100. For space limitation, we will only show the results for SNR = 0.5 and SNR = 2 since the conclusions for the other two values are the same. Moreover, we also consider generating i.i.d. Gaussian error terms, but end up with the same qualitative comparison results.

We generate the error terms for 100 times and then compute the bias^2, the mean squared error (MSE), coverage probability of each individual coefficient and rank the importance of covariates based on the generalized t-statistic.

3.2 Simulation results

3.2.1 Bias-variance trade-off, MSE and coverage

As the authors pointed out, the de-biased Lasso "is not designed for variable selection or the estimation of the coefficients, and the l_2 estimation error of the de-biased Lasso is of order $\sigma^2 p/n$, compared with $\sigma^2 s \log(p)/n$ for the Lasso," where σ^2 is the noise variance, s is number of nonzero coefficients, p is the number of covariates and n is the sample size. However, limited research has been done on the bias-variance trade-off of the de-biased Lasso estimate for each individual regression coefficient.

Figures 1 (fMRI) and 2 (Housing) show the bias^2 (first row), variance (second row) and mean squared error (MSE; third row) of the de-biased Lasso against the LassoOLS. To have a better display, the bias^2, variance and MSE are rescaled by dividing $M = \max_{j=1,\ldots,p}(\beta_j^*)^2$. Each scatter point stands for one individual coefficient, and red points are presented for nonzero coefficients, while green points are presented for zero coefficients. We find that, for the fMRI data set, the de-biased Lasso reduces the bias of the LassoOLS for nonzero coefficients, but increases the bias for zero coefficients. Overall, the LassoOLS has smaller bias^2, which is 4–16% smaller than de-biased Lasso and 84–98% smaller than Lasso; see Figures S3 and S4 in Electronic Supplementary Material for comparison of the LassoOLS to the Lasso. For the housing data set, where the covariates are not normally distributed, even for nonzero coefficients, the de-biased Lasso has larger biases than both the Lasso and the LassoOLS. In terms of variance, Lasso performs the best, which is 35–72% smaller (overall) than the

Author's personal copy

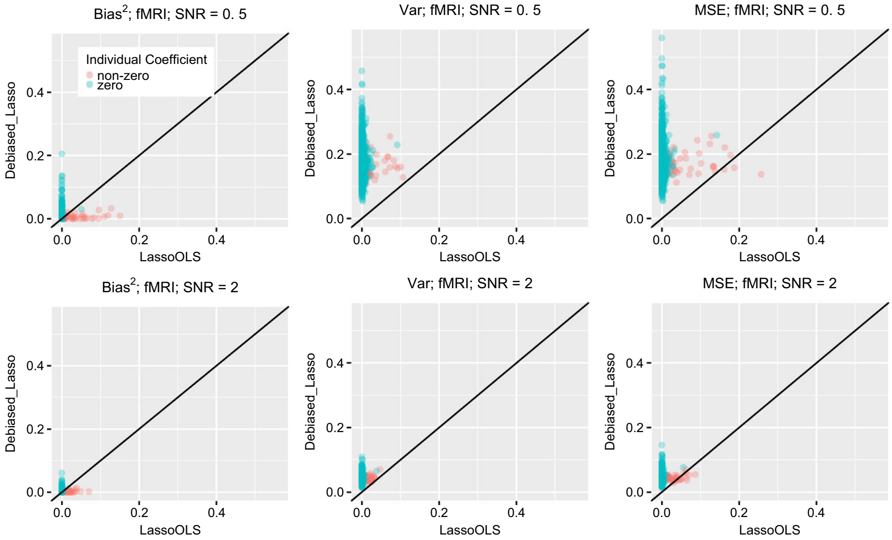

Fig. 1 Relative bias^2 (first column), variance (second column) and MSE (third column) of de-biased Lasso against LassoOLS for each individual coefficient based on fMRI data set. Red points are presented for nonzero coefficients, and green points are presented for zero coefficients. The diagonal solid line corresponds to when two methods perform the same (color figure online)

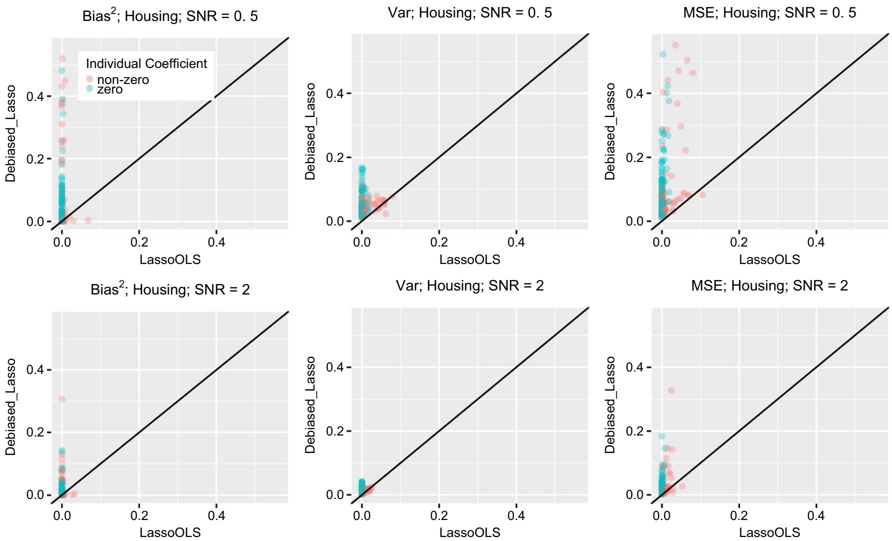

Fig. 2 Relative bias^2 (first column), variance (second column) and MSE (third column) of de-biased Lasso against LassoOLS for each individual coefficient based on housing data set. Red points are presented for nonzero coefficients, and green points are presented for zero coefficients. The diagonal solid line corresponds to when two methods perform the same (color figure online)

LassoOLS and 81–99% smaller than de-biased Lasso. As a result, the Lasso has the best overall MSE: 11–30% better than the LassoOLS and 86–99% better than the de-biased Lasso.

Author's personal copy

Figure 3 displays coverage probability (95% confidence interval) against MSE for each individual coefficient based on fMRI data. Three methods are compared: bootstrapping the de-biased Lasso, bootstrapping the LassoOLS and bootstrapping the Lasso. We see that in order to cover the true coefficients, bootstrapping the de-biased Lasso requires to increase the variances of the estimates, which results in relatively large MSEs, especially for small and zero coefficients. On the other hand, although bootstrapping the LassoOLS (or Lasso) does not produce the correct coverage for nonzero coefficients, it can estimate these coefficients with greater accuracy and is able to correctly identify zero coefficients. Moreover, when SNR increases, its coverage performance improves. Results for Housing data are similar; see Figure S1 in Electronic Supplementary Material.

3.2.2 Ranking and selection of covariates

To compare the ranking performance, we compute the ReDiscovery(k) for $k = 1, \ldots, 20$ and show them in Fig. 4 in the main text and Figure S2 in Electronic Supplementary Material for fMRI data and Housing data, respectively. It is clear that bootstrapping the LassoOLS produces more accurate rankings, especially when SNR is low. Together with its conceptual and computational simplicity, and associated easy accessibility and good reproducibility, for the ranking metric, bootstrapping LassoOLS seems a good choice.

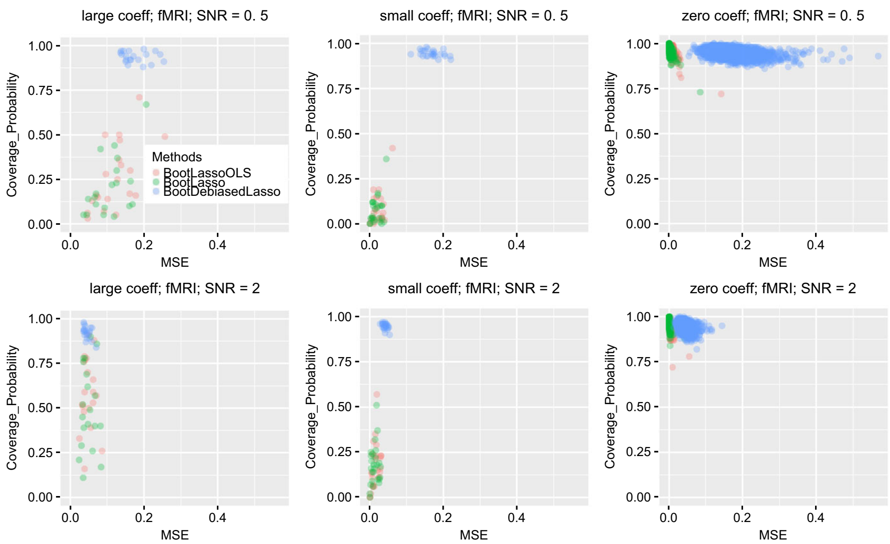

Fig. 3 Coverage probability (95% CI) against MSE for large individual coefficients (the first 20 largest), small individual coefficients (the nonzero coefficients except the first 20 largest) and zero individual coefficients based on fMRI data set. Three methods are compared: bootstrapping the de-biased Lasso (blue points), bootstrapping the LassoOLS (red points) and bootstrapping the Lasso (green points) (color figure online)

Author's personal copy

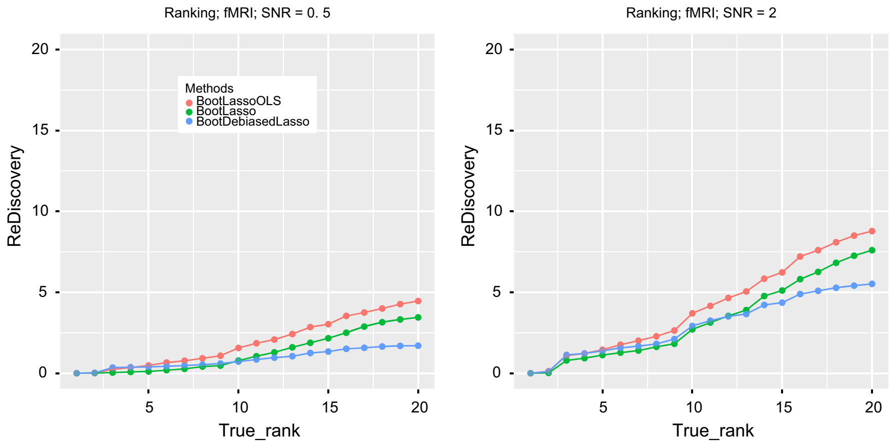

Fig. 4 Comparison of ranking performance of bootstrapping the Lasso (green), bootstrapping the de-biased Lasso (blue) and bootstrapping the LassoOLS (red) based on fMRI data set (color figure online)

4 Conclusion

Estimation and inference in high-dimensional models both appear in decision making as effect size assessment and uncertainty quantification. We made a case for MSE to be an additional metric (to coverage) for evaluation of statistical methods to support decision making. We also made a case for ranking as a useful and relevant metric for evaluating recommendations from modern statistical analysis. Ranking and selecting important covariates/coefficients can be more fundamental and desirable than covering the true coefficient or testing its significance. In two real data-derived simulation studies of sparse linear models with low- and medium-size SNRs, we showed that: (1) the de-biased Lasso is designed to get desired coverage but performs disappointedly for the MSE of estimating the whole coefficients vector or even estimating an individual coefficient because the de-biased Lasso pays a high penalty for variance in order to reduce bias, while the LassoOLS and the Lasso entertain overall smaller MSEs, and (2) bootstrapping the LassoOLS (or Lasso) can work better than bootstrapping the de-biased Lasso in terms of ranking covariates. Together with its conceptual and computational simplicity, we believe bootstrapping the LassoOLS (or Lasso) is a useful and effective method to support decision making. When SNR is not very low (> 2), bootstrapping the LassoOLS seems better; otherwise, bootstrapping the Lasso is preferred.

However, more research both from empirical studies and from theoretical investigations is necessary to fully assess the importance of MSE and ranking as performance metrics on statistical methods for decision-making support. Moreover, the effect of non-Gaussian predictors on the performance of compared inference methods deserves further investigation too.

Acknowledgements We would like to thank the Editor for the invitation to discuss and thank the Gallant Lab at UC Berkeley for providing the fMRI data. We also thank Jasjeet Sekhon for helpful discussions and

Author's personal copy

for comments that clarify the text and thank Rebecca Barter for extremely helpful comments that led to much improvement of the paper.

References

Bahadur RR (1950) On the problem in the theory of k populations. Ann Math Stat 21:362–365

Bahadur RR, Robbins H (1950) The problem of the greater mean. Ann Math Stat 21:469–487

Bechhofer RE (1954) A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann Math Stat 25:16–39

Bechhofer RE, Kiefer J, Sobel M (1968) Sequential identification and ranking procedures. University of Chicago Press, Chicago

Chatterjee A, Lahiri SN (2010) Asymptotic properties of the residual bootstrap for Lasso estimators. Proc Am Math Soc 138:4497–4509

Chatterjee A, Lahiri SN (2011) Bootstrapping Lasso estimators. J Am Stat Assoc 106:608–625

Chatterjee A, Lahiri SN (2013) Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. Ann Stat 41:1232–1259

Dezeure R, Bühlmann P, Meier L, Meinshausen N (2014) High-dimensional inference: confidence intervals, p values and R-software HDI. Statist Sci 30:533–558

Dezeure R, Bühlmann P, Zhang C-H (2017) High-dimensional simultaneous inference with the bootstrap. Test. doi:10.1007/s11749-017-0554-2

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355

Liu H, Yu B (2013) Asymptotic properties of Lasso + mLS and Lasso + ridge in sparse high-dimensional linear regression. Electron J Stat 7:3124–3169

Lu X, Gamst A, Xu R (2009) Rdcurve: a non-parametric method to evaluate the stability of ranking procedures. IEEE/ACM Trans Comput Biol Bioinform. doi:10.1109/TCBB.2008.138

Mukhopadhyay N (2000) A conversation with Milton Sobel. Stat Sci 15(2):168–190