DOI: 10.1111/biom.13020

WILEY Biometrics

# Rejoinder to Discussions on: Model confidence bounds for variable selection

**Yang Li**[1,2] | **Yuetian Luo**[3] | **Davide Ferrari**[4] | **Xiaonan Hu**[5,6] | **Yichen Qin**[7,*]

[1]School of Statistics, Renmin University of China
[2]Center for Applied Statistics, Renmin University of China
[3]Department of Statistics, University of Wisconsin, Madison
[4]Faculty of Economics and Management, Free University of Bozen-Bolzano
[5]Department of Biostatistics, Yale University
[6]School of Mathematical Sciences, University of Chinese Academy of Sciences
[7]Department of Operations, Business Analytics, and Information Systems, University of Cincinnati

**Correspondence**
*Department of Operations, Business Analytics, and Information Systems, University of Cincinnati
yichen.qin@uc.edu

We would like to thank the discussants and editors for their enlightening comments, interesting observations, and recognition of our proposed method. Model selection uncertainty is a fast-growing research topic with constant new results, and it also poses many challenges in theory and computation for analyzing complex data. We think a more comprehensive investigation on model selection uncertainty is in order and hope the current discussions will be helpful for a thorough analysis of such a topic. We look forward to seeing more investigation on model selection uncertainty, and new methods and tools developed to overcome the unsolved challenges. Below, we have organized our discussion in the following sections to address the issues raised by the discussants.

## 1 | CLAESKENS AND JANSEN

The discussants have provided very insightful comments on improvements of the proposed method. The key questions from the discussants are: What if the model space we are working with does not contain the true model? And what if the model space becomes too large, for example for the high-dimensional data and the interaction effects, in which case the cardinality of MCB becomes exponentially large?

It is apparent that we have chosen a very restrictive context, with the true model being assumed to lie in the model space we are working with. We made such a choice because the setting provides the clearest objective and a well-defined framework and concepts. It will be very interesting to see what happens to the MCB when the model is misspecified. In that situation, we would hope that the MCB will target at the projection model, i.e., the projection of true model into the model space we are working on (White, 1982). In other words, this projected true model (or least false model) will be contained by MCB with certain confidence. To investigate such a case, we would need to rely on a couple of tools. First, we would need to formally study the projection model. For example, in what metrics is the projection taken? How would various model selection methods such as Adaptive Lasso and BIC behave in that situation (Lu et al., 2012; Lv and Liu, 2014)? It is not obvious whether all of the methods will lead to the same projection model. Second, we need to identify the appropriate model selection procedure for working with misspecified models. In other words, we would need to first establish the "selection consistency" under the misspecified model settings for various model selection procedures, and then extend these results to the MCB. There have been many successful investigations in such a setting, so we anticipate that this is a great topic for future research.

Meanwhile, the discussants have also raised another critical issue of the proposed method about the increased model complexity. In our article, we mainly focus on the low dimension case. However, when dealing with high-dimensional data, the model selection uncertainty will inevitably increase, leading to a much larger cardinality of the MCB. This is because, as the dimension increases, the model space increases exponentially; for example, $p$ covariates will result in $2^p$ possible models without even considering interaction effects. Only when the true model is easy to select and the signal is strong will the MCB still be able to maintain low cardinality.

In the case of high-dimensional data, the proposed MCB may not be able to offer sufficient information about the model selection uncertainty. However, we can improve it by adapting more structures. Currently, MCB is equipped with one lower bound and one upper bound. Instead of identifying only one lower bound, we could identify multiple lower bounds (Ferrari and Yang, 2015). In other words, as the dimension increases, we could add more structures to the MCB and find a compromise between MCS (Model Confidence Set) and MCB. Note that MCB is essentially a special case of MCS, but MCS is much harder to interpret. Balancing between MCB and a general MCS essentially means finding a sweet spot with accurate measurement of model selection uncertainty and maximum interpretability. This also requires more investigation on the distribution of model selection results.

## 2 | LEEB, PÖTSCHER, AND KIVARANOVIC

We really appreciate the interesting observations and insightful comments from the discussants, especially the strengthening of our theorem. Our method can be considered as an extension of the traditional confidence interval to the model selection context, that is, treating the model selection procedure as an "estimator". Since the lower bound model (LBM) and upper bound model (UBM) in a MCB are derived from the bootstrap models, as the sample size increases, the LBM and UBM converge to a single model, the true model. This is similar to the case of confidence interval for the population parameter in the sense that the two boundaries of the interval converge to the true parameter. However, in contrast to the traditional confidence interval case where the true and estimated parameters belong to a Euclidean space, there are only a finite number of possible models that could be selected. This nature of model selection gives MCB several special properties. As the sample size increases, the probabilities of LBM being the true model and UBM being the true model are converging to 1. Therefore, the coverage probability of MCB will also converge to 1 in the limit, which is different from the traditional confidence interval, whose coverage probability converges to $1-\alpha$. Meanwhile, for an MCB at a finite sample size, it is very hard to determine the exact coverage probability. Although we have focused on the fixed dimension setting where $n$ increases while $p$ is fixed, it will be very interesting to investigate the properties of the coverage probability in the case of growing dimension.

We also appreciate the discussants for pointing out the drawback of the proposed method in the varying-parameter setting. It is indeed important to provide a uniform asymptotic analysis. As the discussants explained in their earlier work, many penalized likelihood estimation methods may suffer in this weak effect setting. Since MCB is often based on penalized likelihood estimation, it is subject to the same shortcoming. In fact, similar to the confidence interval constructed by bootstrapping Adaptive Lasso or other methods, our MCB needs the so-called "beta-min" condition, which assumes that the nonzero coefficients are not too small. However, it is worth mentioning that MCB is not restricted to being tied with penalized estimation methods, but also capable of incorporating many other model selection methods, such as BIC. Therefore, it would be interesting to see if these methods will survive such a setting.

Providing uniform asymptotic results under the varying-parameter setting is indeed important for constructing a hypothesis test or confidence interval when focusing on the magnitude of the parameters (Dezeure et al., 2017), but the fixed-parameter setting is still an important case that requires thorough investigation. Recent research has also been conducted on the fixed-parameter setting, such as Chatterjee and Lahiri (2011, 2013) and Liu and Yu (2013, 2017). Our article focuses on the fixed parameter setting with asymptotic properties established by letting the sample size go to infinity. When fixing the sample size as the discussants did, varying the parameters to mimic the weak effect will often lead to unsatisfactory results. In practice, the effect size and level of significance are strongly domain-dependent, as illustrated in Liu and Yu (2017). Dezeure et al. (2017) recently proposed a new confidence interval by bootstrapping de-biased Lasso, which can overcome the challenge of the weak effect. However, due to the properties of de-biased Lasso, their confidence interval sometimes has a large width for small coefficients. On the other hand, the confidence interval constructed by bootstrapping Lasso or Adaptive Lasso is conceptually simpler and computationally easier, and often provides relatively short intervals. Even though the confidence interval from bootstrapping Lasso for small coefficients leads to near zero coverage and sometimes zero length, such a case in real world problems where such coefficients are the goal of analysis is not commonly encountered. In addition, Liu and Yu (2017) showed that bootstrapping Lasso and de-biased Lasso generate a comparable performance in simulation. In fact, under weak effect with small coefficients, neither bootstrapping de-biased Lasso nor Lasso dominates the other. Bootstrapping de-biased Lasso gives wider confidence intervals and better coverage rate, while bootstrapping the Lasso gives small confidence intervals or zero confidence intervals. In addition, note that in our proposed method, we are not targeting at the parameter confidence interval, instead, at model confidence bound which is discrete as we explained at the beginning of this section.

**TABLE 1** Comparison of coverage rates of MCB constructed by different model selection methods.

| Width | BIC | SCAD | Lasso |
|---|---|---|---|
| 1 | 0.12 | 0.04 | 0.50 |
| 2 | 0.20 | 0.07 | 0.72 |
| 3 | 0.27 | 0.08 | 0.79 |
| 4 | 0.34 | 0.11 | 0.86 |
| 5 | 0.43 | 0.12 | 0.90 |
| 6 | 0.51 | 0.16 | 0.93 |
| 7 | 0.60 | 0.20 | 0.95 |
| 8 | 0.69 | 0.25 | 0.95 |
| 9 | 0.79 | 0.41 | 0.96 |
| 10 | 1.00 | 1.00 | 1.00 |
| 11 | 1.00 | 1.00 | 1.00 |
| 12 | 1.00 | 1.00 | 1.00 |
| 13 | 1.00 | 1.00 | 1.00 |
| 14 | 1.00 | 1.00 | 1.00 |
| 15 | 1.00 | 1.00 | 1.00 |

Additionally, we replicate a small simulation study under the same setting as the discussants, i.e., finite sample with varying-parameter, where $p = 15$ and $p^* = 6$. Since the definition of MCB includes a pair of upper and lower bounds with the shortest width among MCBs satisfying any given confidence level $1 - \alpha$, three interesting conclusions can be implied from Table 1. First, the MCB approach performs well in numerical study even for varying-parameter with finite-sample and more theoretical analysis is needed in future work. Second, the optimal width should be searched strictly under the definition instead of arbitrarily using $p^*$. Third, different model selection methods (BIC, SCAD, and Lasso) may have different MCB widths which makes the proposed method feasible for the selection of model selection methods.

In the end, we would also like to point out that MCB represents a class or template of algorithms in the sense that MCB can incorporate many existing model selection methods, and we need to work out the instantiation for each particular case, i.e., each model selection method. The advantage of MCB is that it can be coupled with many model selection methods; therefore, it is extremely flexible and can be adopted to many contexts as long as the model selection method applies, hence providing a platform to compare many model selection methods. On the other hand, many of the properties of MCB depend heavily on the model selection method at hand. For example, if the model selection employed by MCB does not perform well, MCB will also suffer. Many discussions on MCB are essentially rooted in the properties of the model selection methods. Fortunately, model selection methods are under rapid development, and many current open questions will likely be resolved by advances in this area. We look forward to seeing more exciting methods in the future to address these challenges.

## ACKNOWLEDGEMENTS

This work is partially supported by the Research Fund (15XNI011) of Renmin University of China.

## REFERENCES

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *J Am Stat Assoc* 106, 608–625.

Chatterjee, A. and Lahiri, S. N. (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., and Zhang, C.-H. (2017).High-dimensional simultaneous inference with the bootstrap. *TEST* 26, 685–719.

Ferrari, D. and Yang, Y. (2015).Confidence sets for model selection by F-testing. *Statistica Sin* 25, 1637–1658.

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

Liu, H. and Yu, B. (2017).Comments on: High-dimensional simultaneous inference with the bootstrap. *TEST*26, 740–750.

Lu, W., Goldberg, Y., and Fine, J. P. (2012).On the robustness of the adaptive lasso to model misspecification. *Biometrika* 99, 717–731.

Lv, J. and Liu, J. S. (2014).Model selection principles in misspecified models. *J R Stat Soc Ser B (Stat Methodol)* 76, 141–167.

White, H. (1982).Maximum likelihood estimation of misspecified models. *Econometrica* 101, 1–25.