Supplementary Material for “Markov Neighborhood Regression for Statistical Inference of High-Dimensional Generalized Linear Models”

Lizhe Sun and Faming Liang *

May 22, 2022

The supplementary material is organized as follows. Section 1 provides a selective review for the existing high-dimensional inference methods, where the Markov neighborhood regression (MNR) method is excluded as it has been described in Section 2 of the main text. Section 2 demonstrates the robustness of the MNR method with respect to the dependence structure among explanatory variables. Section 3 presents some cross-validation results for the CCLE data analyzed in the main text.

1 A Selective Review of the Existing High-Dimensional Inference Methods

1.1 Desparsified Lasso

The desparsified Lasso method[1] is essentially the same as the one developed in Ref[2] and Ref[3]. For the Gaussian high-dimensional linear regression

\[ Y = X\beta + \epsilon, \]

[1] To whom correspondence should be addressed: Faming Liang. F. Liang is Professor (email: fmliang@purdue.edu), L. Sun is Postdoc, Department of Statistics, Purdue University, West Lafayette, IN 47907.
desparsified Lasso defines a bias-corrected estimator

\[
\hat{\beta}_{bc} = \hat{\beta}_{Lasso} + \hat{\Theta} X^T (y - X \hat{\beta}_{Lasso}) / n,
\]

(S1)

where \( \hat{\beta}_{Lasso} \) is the original Lasso estimator, and \( \hat{\Theta} \) is an approximator to the inverse of \( \hat{\Sigma} = X^T X / n \). From (S1), one can obtain

\[
\sqrt{n}(\hat{\beta}_{bc} - \beta) = \hat{\Theta} X^T \epsilon / \sqrt{n} + \sqrt{n}(I_p - \hat{\Theta} \hat{\Sigma})(\hat{\beta}_{Lasso} - \beta) := \hat{\Theta} X^T \epsilon / \sqrt{n} + \Delta_n,
\]

(S2)

where \( I_p \) denotes the \( p \times p \) identity matrix, and \( \Delta_n \) is called the error term. With an appropriate estimator \( \hat{\Theta} \), e.g., the one obtained via nodewise regression\(^4\), it can be shown that \( \| \Delta_n \|_{\infty} = o_p(1) \), which implies that \( \sqrt{n}(\hat{\beta}_{bc} - \beta) \) shares the same asymptotic distribution with \( \hat{\Theta} X^T \epsilon / \sqrt{n} \). Further, to calculate confidence intervals for \( \beta \), one needs to approximate the distribution of \( \hat{\Theta} X^T \epsilon / \sqrt{n} \). For example, Ref\(^3\) proposed to approximate it by \( N(0, \hat{\sigma}^2 \hat{\Theta} \hat{\Sigma} \hat{\Theta}^T) \), where \( \hat{\sigma}^2 \) is a consistent estimator of \( \sigma^2 \); and Ref\(^5\) proposed to approximate it using multiplier bootstrap.

1.2 Ridge Projection

The ridge projection method\(^6\) can be viewed as a direct extension of the low-dimensional ridge regression to the high-dimensional case. The bias of the ridge estimator has been assessed and corrected, the distribution of the ridge estimator has been derived and approximated, and thus it can be used for high-dimensional inference.

1.3 Multi Sample-Splitting Method

The multi sample-splitting method\(^7\) works in the following procedure: Splitting the samples into two subsets equally, using the first half of samples for variable selection and using the second half of samples with the selected variables for calculating \( p \)-values; repeating this process for many times; and aggregating the \( p \)-values for statistical inference. The confidence intervals can be constructed based on their duality with the \( p \)-values. The idea about sample-splitting and subsequent statistical inference has also been implicitly contained in Ref\(^8\). The multi sample-splitting method is very general and can be applied to many different types
of models. We have conducted some numerical studies with this method, and found that it often performs less well compared to desparsified Lasso and the proposed MNR method.

1.4 Other Methods

The other methods include residual-type bootstrapping\textsuperscript{9,10}, covariance test\textsuperscript{11}, and group-bound\textsuperscript{12}. A problem with the residual-type bootstrapping method is the super-efficiency phenomenon; that is, the confidence interval of a zero regression coefficient might be the singleton \{0\}. The covariance test method relies on the solution path of the Lasso and is much related to the post-selection inference methods\textsuperscript{13–16}. The group-bound method provides a good treatment for highly correlated variables, but has often a weak power in detecting the effect of individual variables. Recently, some methods based on the idea of estimating a low-dimensional component of a high-dimensional model have also been proposed, see e.g. Ref\textsuperscript{17} and Ref\textsuperscript{18}.

2 Robustness of the MNR Method

This section demonstrates the robustness of the MNR method with respect to the dependence structure among the explanatory variables.

2.1 Non-Sparse Adjacency Matrix

We generated 500 data sets from a linear regression model with \( p = 500, n = 300 \) and \( \sigma^2 = 1 \). The true regression coefficients were given by \((\beta^*_0, \beta^*_1, \beta^*_2, \cdots, \beta^*_p) = (1, 2, 2.5, 3, 3.5, 4, 0, \ldots, 0)\). The features were generated using the same method as described in Section 4.1 of main text except that the adjacency matrix was given by

\[
E_{i,j} = \begin{cases} 
1, & \text{if } j > i, i = 1, \cdots, (j - 1), \\
0, & \text{otherwise}, 
\end{cases}
\]  
(S3)
and the equation (14) was changed as follows:

\[ Z_j = \sum_{i=1}^{j-1} E_{ij} X_i / j, \]

\[ X_j = \begin{cases} 
Z_j + \epsilon, & \text{if } X_j \text{ is continuous,} \\
\text{Binomial}(1, \frac{\exp(Z_j)}{1+\exp(Z_j)}) & \text{if } X_j \text{ is binary,}
\end{cases} \]

for \( j = 2, 3, \ldots, p \), where half of the features are continuous and the other half of the features are discrete. Note that the adjacency matrix given in (S3) violates the sparsity condition assumed for the graphical model.

Table S1: Coverage rates and widths of the 95% confidence intervals produced by desparsified Lasso, ridge projection and MNR for the simulated linear regression data with a non-sparse adjacency matrix. Refer to Table 1 of the main text for the notation.

| Measure | Desparsified-Lasso | Ridge Projection | MNR |
|---------|--------------------|------------------|-----|
| Coverage signal | 0.815 (0.017) | 0.974 (0.007) | 0.945 (0.010) |
|          noise | 0.928 (0.012) | 0.979 (0.006) | 0.950 (0.010) |
| Width signal | 0.357 (0.006) | 0.652 (0.010) | 0.364 (0.005) |
|          noise | 0.371 (0.006) | 0.645 (0.010) | 0.354 (0.006) |

The desparsified Lasso, ridge projection, and MNR methods were applied to the datasets. The MNR method was implemented using the package SIS\textsuperscript{19}, where SIS-MCP was used for both variable selection and nodewise regression. The other two methods have been implemented in the R package hdi\textsuperscript{20}. The results are summarized in Table S1. The comparison indicates that the MNR method is robust to the violation of the sparsity condition imposed on the Markov network, and it significantly outperforms other methods in coverage probability.

The multi sample-splitting method has also been tried for the linear regression data sets. We found that it is super-efficient in converge probability for both zero and nonzero regression coefficients. In addition, the width of the confidence interval for the zero regres-
sion coefficients can be extremely wide. In summary, its performance is not comparable to desparsified Lasso and ridge projection, and thus not included in Table S1.

2.2 Random Adjacency Matrix

In this section, we present one more simulation case to demonstrate the robustness of the MNR method. Also, we generated 500 data sets from a linear regression model with $p = 500$, $n = 300$ and $\sigma^2 = 1$. The true regression coefficients were given by $(\beta_0^*, \beta_1^*, \beta_2^*, \ldots \beta_p^*) = (1, 2, 2.5, 3, 3.5, 4, 0, \ldots, 0)$. The variables were generated as described in Section 4.1 of main text except that the adjacency matrix were random:

$$E_{i,j} = \begin{cases} 1, & \text{random distributed for } j > i, i = 1, \ldots, (j - 1), \\ 0, & \text{otherwise}, \end{cases}$$

for which we set the sparsity level to 0.01. And we set the parameter $\rho = 0.5$ in this case. The results are summarized in Table S2.

Table S2: Coverage rates and widths of the 95% confidence intervals produced by desparsified Lasso, ridge projection and MNR for the simulated linear regression data with a random adjacency matrix. Refer to Table 1 of the main text for the notation.

| Measure | — | Desparsified-Lasso | Ridge Projection | MNR |
|---------|---|---------------------|------------------|-----|
| Coverage | signal | 0.945 (0.010) | 0.976 (0.007) | 0.953 (0.009) |
|          | noise  | 0.953 (0.009) | 0.982 (0.006) | 0.950 (0.010) |
| Width   | signal | 0.389 (0.005) | 0.713 (0.009) | 0.384 (0.004) |
|          | noise  | 0.382 (0.007) | 0.699 (0.012) | 0.381 (0.006) |
3 Identification of Drug Sensitive Genes and Mutations

This section presents some cross-validation results in Table S3 for the CCLE data analyzed in the main manuscript.

Table S3: Comparison of MNR with SIS-SCAD, SIS-MCP and SIS-Lasso in gene/mutation selection for 24 drugs: “MSPE” refers to the mean squared prediction error and “Size” refers to the number of selected gene/mutations, which are reported as the average over 5-fold results with the standard deviation given in the parentheses; $p$-value is calculated with a pair-$t$ test for the hypotheses $H_0$: MSPE(MNR)=MSPE(m) versus $H_1$: MSPE(MNR) $\neq$ MSPE(m) based on the cross-validation results, where $m \in \{\text{SIS-SCAD, SIS-MCP, SIS-Lasso}\}$ and MSPE(m) denotes the MSPE produced by the method m.

| Drug       | -     | SIS-SCAD        | SIS-MCP        | SIS-Lasso      | MNR       |
|------------|-------|-----------------|----------------|----------------|-----------|
|            | MSPE  | 0.876(0.16)     | 0.893(0.14)    | 0.987(0.10)    | 0.976(0.11)|
| 17-AAG     | $p$-value | 0.281        | 0.334        | 0.877        | -         |
|            | Size   | 20(11.47)      | 16.2(3.49)    | 7.8(11.03)    | 1.2(0.45) |
| AEW541     | MSPE  | 0.892(0.05)     | 0.895(0.06)    | 0.916(0.05)    | 1.022(0.10)|
|            | $p$-value | 0.037        | 0.043        | 0.076        | -         |
|            | Size   | 17.8(3.56)      | 12.2(3.27)     | 6.2(7.05)     | 1.2(0.45) |
| AZD0530    | MSPE  | 1.130(0.36)     | 1.138(0.35)    | 1.089(0.35)    | 1.029(0.34)|
|            | $p$-value | 0.662        | 0.630        | 0.790        | -         |
|            | Size   | 12.8(11.39)     | 14(2.12)      | 11.2(11.26)   | 1.2(0.45) |
| AZD6244    | MSPE  | 0.682(0.08)     | 0.650(0.05)    | 0.677(0.07)    | 0.972(0.23)|
|            | $p$-value | 0.047        | 0.035        | 0.045        | -         |
|            | Size   | 12.8(3.27)      | 7.2(2.28)     | 9.6(3.36)     | 1.2(0.45) |
| Erlotinib  | MSPE  | 0.982(0.38)     | 0.969(0.40)    | 0.948(0.44)    | 1.104(0.52)|
|            | $p$-value | 0.683        | 0.657        | 0.620        | -         |
|            | Size   | 11.2(1.92)      | 6.6(2.51)     | 10(5.43)      | 2.2(1.79) |
|            | MSPE  | 0.552(0.08)     | 0.559(0.09)    | 0.542(0.09)    | 0.750(0.07)|
| Drug       | p-value | Size    | MSPE     |
|------------|---------|---------|----------|
| Irinotecan | 0.003   | 6.6(0.89) | 0.788(0.21) |
|           | 0.006   | 3.8(0.84) | 0.815(0.22) |
|           | 0.004   | 9.8(3.03) | 0.767(0.19) |
|           | -       | 1.0(0.0)  | 0.974(0.28) |
| L-685458  | 0.271   | 8.6(2.51) | 0.763(0.15) |
|           | 0.349   | 5.0(1.87) | 0.772(0.15) |
|           | 0.219   | 14.0(1.0) | 0.760(0.16) |
|           | -       | 2.0(2.24) | 0.912(0.21) |
| Lapatinib | 0.233   | 9.2(2.86) | 1.154(0.21) |
|           | 0.258   | 5.0(3.46) | 1.275(0.38) |
|           | 0.230   | 11.2(2.86) | 1.084(0.31) |
| LBW242    | -       | 1.6(0.55) | 1.182(0.63) |
| Nilotinib | 0.567   | 10.2(2.59) | 0.952(0.59) |
|           | 0.564   | 6.6(1.82) | 0.946(0.62) |
|           | 0.423   | 13.6(2.86) | 0.878(0.51) |
| Nutlin-3  | -       | 3.0(2.35) | 1.182(0.63) |
| Paclitaxel| 0.000   | 15.4(1.82) | 0.988(0.14) |
|           | 0.000   | 9.8(1.64) | 1.009(0.14) |
|           | 0.000   | 17.0(3.74) | 1.041(0.21) |
|           | -       | 1.0(0.0)  | 1.013(0.15) |
| Panobinostat| 0.002  | 7.2(1.10) | 0.720(0.07) |
|           | 0.005   | 3.6(0.89) | 0.716(0.08) |
|           | 0.001   | 13(3.16)  | 0.724(0.08) |
| PD-0325901| -       | 1.0(0.0)  | 1.047(0.08) |
| PD-0332991| 0.027   | 14.0(2.0) | 0.613(0.06) |
|           | 0.028   | 7.6(0.55) | 0.613(0.05) |
|           | 0.022   | 13.2(5.89) | 0.588(0.05) |
|           | -       | 1.0(0.0)  | 0.956(0.23) |
| PD-0332991| 0.013   | 9.0(2.0)  | 0.753(0.14) |
|           | 0.021   | 5.4(0.89) | 0.782(0.14) |
|           | 0.012   | 12.2(2.59) | 0.756(0.13) |
|           | -       | 1.0(0.0)  | 1.054(0.16)|
| Drug     | MSPE       | 0.948(0.19) | 0.923(0.18) | 0.891(0.17) | 0.997(0.19) |
|----------|------------|-------------|-------------|-------------|-------------|
|          | p-value    | 0.696       | 0.540       | 0.373       | -           |
|          | Size       | 14.2(1.12)  | 10.4(2.51)  | 11.4(4.95)  | 1.8(1.30)   |
| PF2341066| MSPE       | 1.001(0.27) | 0.984(0.24) | 0.977(0.27) | 1.037(0.25) |
|          | p-value    | 0.83        | 0.737       | 0.727       | -           |
|          | Size       | 12.6(3.65)  | 7.8(1.92)   | 7.2(5.36)   | 2.2(1.79)   |
| PHA-665752| MSPE       | 0.908(0.28) | 0.889(0.27) | 0.866(0.23) | 0.646(0.12) |
|          | p-value    | 0.106       | 0.12        | 0.11        | -           |
|          | Size       | 9.8(5.12)   | 5.4(2.79)   | 10.2(5.59)  | 3.2(3.8)    |
| PLX4720  | MSPE       | 0.942(0.30) | 0.901(0.22) | 0.988(0.27) | 0.996(0.25) |
|          | p-value    | 0.768       | 0.543       | 0.962       | -           |
|          | Size       | 13.2(10.89) | 13.8(1.30)  | 0.6(1.34)   | 1.0(0.0)    |
| RAF265   | MSPE       | 1.136(0.32) | 0.979(0.28) | 0.966(0.29) | 1.039(0.39) |
|          | p-value    | 0.679       | 0.786       | 0.746       | -           |
|          | Size       | 12.2(2.77)  | 7.0(1.22)   | 12.8(4.55)  | 1.0(0.0)    |
| Sorafenib| MSPE       | 0.909(0.14) | 0.864(0.14) | 0.846(0.14) | 1.006(0.20) |
|          | p-value    | 0.404       | 0.237       | 0.184       | -           |
|          | Size       | 11.4(7.13)  | 10(2.65)    | 7.6(4.56)   | 1.0(0.0)    |
| TAE684   | MSPE       | 0.985(0.15) | 0.960(0.13) | 0.953(0.17) | 1.059(0.22) |
|          | p-value    | 0.554       | 0.414       | 0.418       | -           |
|          | Size       | 11.8(2.49)  | 6.8(3.03)   | 11.8(5.59)  | 1.4(0.89)   |
| TKI258   | MSPE       | 0.599(0.10) | 0.601(0.12) | 0.588(0.12) | 0.801(0.18) |
|          | p-value    | 0.07        | 0.08        | 0.06        | -           |
|          | Size       | 13.6(1.14)  | 7.4(1.82)   | 18.8(0.84)  | 1.0(0.0)    |
| Topotecan| MSPE       | 0.977(0.15) | 0.971(0.16) | 0.995(0.14) | 1.017(0.15) |
|          | p-value    | 0.685       | 0.649       | 0.819       | -           |
|          | Size       | 8.8(8.7)    | 12(2.12)    | 1.6(2.61)   | 1.0(0.0)    |
| ZD-6474  | MSPE       | 0.977(0.15) | 0.971(0.16) | 0.995(0.14) | 1.017(0.15) |
|          | p-value    | 0.685       | 0.649       | 0.819       | -           |
|          | Size       | 8.8(8.7)    | 12(2.12)    | 1.6(2.61)   | 1.0(0.0)    |
References

[1] van de Geer S, Bühlmann P, Ritov Y, Dezeure R. On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics*. 2014;42:1166-1202.

[2] Zhang CH, Zhang SS. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2014;76:217–242.

[3] Javanmard A, Montanari A. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*. 2014;15:2869-2909.

[4] Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*. 2006;34:1436–1462.

[5] Zhang X, Cheng G. Simultaneous Inference for High-Dimensional Linear Models. *Journal of the American Statistical Association*. 2017;112:757-768.

[6] Bühlmann P. Statistical significance in high-dimensional linear models. *Bernoulli*. 2013;19:1212-1242.

[7] Meinshausen N, Meier L, Bühlmann P. $p$-values for high-dimensional regression. *Journal of the American Statistical Association*. 2009;104:1671-1681.

[8] Wasserman L., Roeder K.. High-Dimensional variable selection. *Annals of Statistics*. 2009;37:2178-2201.

[9] Chatterjee A, Lahiri SN. Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Annals of Statistics*. 2013;41:1232-1259.

[10] Liu Y, Yu B. Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics*. 2013;7:3124-3169.

[11] Lockhart R., Taylor J., Tibshirani R.J., Tibshirani B.. A significant test for the Lasso. *Annals of Statistics*. 2014;42:413-468.
[12] Meinshausen N. Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *Journal of the Royal Statistical Society, Series B.* 2015;77:923-945.

[13] Berk RA, Brown LD, Buja A, Zhang K, Zhao LH. Valid post-selection inference. *Annals of Statistics.* 2013;41:802-837.

[14] Lee JD, Sun DL, Sun Y, Taylor JE. Exact post-selection inference, with application to the lasso. *Annals of Statistics.* 2016;44:907-927.

[15] Tibshirani RJ, Taylor J, Lockhart R, Tibshirani R. Exact Post-Selection Inference for Sequential Regression Procedures. *Journal of the American Statistical Association.* 2016;111:600-620.

[16] Fithian W, Sun D, Taylor J. Optimal Inference After Model Selection. *ArXiv:1410.2597.* 2014.

[17] Belloni A, Chernozhukov V, Kato K. Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika.* 2015;102:77-94.

[18] Yang Y. Statistical inference for high dimensional regression via constrained Lasso. *arXiv:1704.05098v1.* 2017.

[19] Saldana DF, Feng Y. SIS: An R Package for Sure Independence Screening in Ultrahigh-Dimensional Statistical Models. *Journal of Statistical Software.* 2018;83:1–25.

[20] Dezeure R, Bühlmann P, Meier L, Meinshausen N. High-dimensional inference: Confidence intervals, p-values and r-software hdi. *Statistical science.* 2015;30:533-558.