Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL TO "A LAVA ATTACK ON THE RECOVERY OF SUMS OF DENSE AND SPARSE SIGNALS"

By Victor Chernozhukov, Christian Hansen and Yuan Liao

Massachusetts Institute of Technology, University of Chicago and University of Maryland

This supplementary material contains additional simulation results and omitted proofs for "A lava attack on the recovery of sums of sparse and dense signals".

CONTENTS

1. Additional simulations.

1.1. *$p = n/2$ and $p = n$ Simulations.* To supplement the simulation results in the main paper, here we present additional simulation results with $p = n/2$ and $p = n$. We consider a fixed design setting analogous to those considered in the main text in the $p = n/2$ setting, and we consider a random design in the $p = n$ setting. We set $n = 100$ in the simulation experiments.

The results in the fixed design cases are reported in Figures 1 and 2, and the results in the random design setting are provided in Figure 3. In all cases, we generate covariates by drawing the rows of X independently from a mean zero multivariate normal with covariance matrix Σ. For the fixed design settings, we consider an independent design, $\Sigma = I$ and a design with a factor covariance structure with $\Sigma = LL' + I$ where the rows of L are independently generated from $N(0, I_3)$. In the latter case, the columns of X depend on three common factors. In these two fixed design cases, we generate the design matrix X once and fix it across simulation replications. In the random design case, we again draw rows of X independently from a mean zero multivariate normal with covariance

AMS 2000 subject classifications: Primary 62J07; secondary 62J05
Keywords and phrases: high-dimensional models, penalization, shrinkage, non-sparse signal recovery

matrix $\Sigma = I$, but resample X in each simulation replication. We compare ols, lasso, ridge, elastic net, lava, and post-lava in the $p = n/2$ simulations, and we use lasso, ridge, elastic net, lava, and post-lava in the $p = n$ simulations. For lasso, ridge, elastic net, lava, and post-lava, we report one set of results based on using SURE minimization for penalty parameter selection and a different set of results based on tuning parameters selected by 5-fold cross-validation. The comparisons are similar to those in the main paper, with lava and post-lava dominating the other procedures.

1.2. *Highly correlated designs.* Consider a highly correlated fixed design, where the regressors are generated as follows:

$$\begin{aligned} X_j &= Z + 0.5\epsilon_j, \quad j = 1, ..., 10 \\ X_j &= \epsilon_j, \quad j = 11, ..., p, \end{aligned}$$

where $\epsilon_j \sim N(0, 1)$, and $Z \sim U(0, 5)$ is independent of ϵ_j. The first ten regressors depend on a common component Z, with an equi-mutual-correlation 0.89. The coefficients are generated as before, and we set $n = 100$, $p = 2n$. The tuning parameters for each method are selected by 5-fold cross-validation. The results are reported in Figure 4. The comparisons are similar to those in the main paper, with lava and post-lava dominating the other procedures.

1.3. *Gaussian sequence model with very high dimension.* We consider a Gaussian sequence model

$$Z = \theta + \epsilon, \quad \epsilon \sim N_p(0, \sigma^2 I_p), \quad \sigma^2 = \frac{1}{n},$$

with $p = \dim(\theta) = 10,000$ and $n = [6\log(p)] = 55$. We set $\theta = (3, q, ..., q)'$ and use the "canonical plug-in" and optimal (risk minimizing) choices for the tuning parameters. Figure 5 compares risks of lava, post-lava, lasso, and ridge estimators as functions of the size of the small coefficients q. The performance is qualitatively similar to that in the main paper even when p is very large, with lava and post-lava dominating the other procedures.

1.4. *Feasible SURE.* Figure 6 plots the results of a feasible SURE minimization procedure where we conservatively estimate σ_u^2 in the $p = 2n$ setting with random design. As before in the random design, we draw rows of X independently from a mean zero multivariate normal with covariance matrix $\Sigma = I$ and resample the design matrix at each simulation replication. The conservative estimate of σ_u^2 is then constructed by taking the estimate of σ_u^2 obtained by using the coefficients obtained from lasso with tuning parameter selected by minimizing 5-fold cross-validation. Estimation of all methods then proceeds by taking the SURE minimizing values of tuning parameters given this conservative estimate of σ_u^2. The results are qualitatively similar to those given for the case of the known error variance σ_u^2 in the sense that lava and post-lava dominate all other procedures.

1.5. *Out-of-sample prediction comparison.* We also compare the out-of-sample prediction performances of the competing estimators within the random design setting. Specifically, we respectively generate training data $\{X_{tr,i}, Y_{tr,i}\}_{i\leq n}$, and testing data $\{X_{test,i}, Y_{test,i}\}_{i\leq m}$ from a linear model $Y_i = X_i^T\theta + U_i$, where $\theta = \delta + q(0, 1...., 1)^T$ is as in the previous simulations. The estimator $\widehat{\theta}_e$ is then obtained using the training data, whose out-of-sample prediction performance is then evaluated using the testing data through the out-of-sample percentage of variation explained (PVE):

$$\mathrm{PVE}(e) = 1 - \frac{\sum_{i=1}^m (Y_{test,i} - X_{test,i}^T\widehat{\theta}_e)^2}{\sum_{i=1}^m (Y_{test,i} - \bar{Y}_{test})^2}, \quad \bar{Y}_{test} = \frac{1}{m}\sum_{i=1}^m Y_{test,i}.$$

We set $n = 100, m = 50$. Figure 7 plots the averaged PVE(e) from 100 replications as a function of the size of small coefficients q. The tuning parameters are chosen via minimizing 5-fold cross validation.

2. Omitted proofs.

2.1. *A simple inequality.* We now prove $\|\mathsf{K}_{\lambda_2} X\beta_0\|_2^2 \leq \|\mathsf{K}_{\lambda_2}^{1/2} X\beta_0\|_2^2 \|\mathsf{K}_{\lambda_2}\| \leq C\|\mathsf{K}_{\lambda_2} X\beta_0\|_2^2$ for some $C > 0$ uniformly in p, n. This implies (3.15) in the main paper.

Let $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ respectively denote the operator of the maximum and minimum eigenvalue. In fact, $\|\mathsf{K}_{\lambda_2} X\beta_0\|_2^2 = \|\mathsf{K}_{\lambda_2}^{1/2}\mathsf{K}_{\lambda_2}^{1/2} X\beta_0\|_2^2 \leq \|\mathsf{K}_{\lambda_2}^{1/2}\|^2 \|\mathsf{K}_{\lambda_2}^{1/2} X\beta_0\|_2^2 = \|\mathsf{K}_{\lambda_2}^{1/2} X\beta_0\|_2^2 \|\mathsf{K}_{\lambda_2}\|$, where the last equality is due to $\|\mathsf{K}_{\lambda_2}^{1/2}\|^2 = \lambda_{\max}(\mathsf{K}_{\lambda_2}^{1/2}\mathsf{K}_{\lambda_2}^{1/2}) = \lambda_{\max}(\mathsf{K}_{\lambda_2}) = \|\mathsf{K}_{\lambda_2}\|$,

Since $\|\mathsf{K}_{\lambda_2}\| \leq 1$,

$$\begin{aligned}\|\mathsf{K}_{\lambda_2}^{1/2} X\beta_0\|_2^2 \|\mathsf{K}_{\lambda_2}\| &= \|\mathsf{K}_{\lambda_2}^{-1/2}\mathsf{K}_{\lambda_2} X\beta_0\|_2^2 \|\mathsf{K}_{\lambda_2}\| \leq \|\mathsf{K}_{\lambda_2} X\beta_0\|_2^2 \|\|\mathsf{K}_{\lambda_2}^{-1/2}\|^2 \|\mathsf{K}_{\lambda_2}\| \\ &= \|\mathsf{K}_{\lambda_2} X\beta_0\|_2^2 \|\mathsf{K}_{\lambda_2}^{-1}\|\|\mathsf{K}_{\lambda_2}\|\end{aligned}$$

Let $\{v_j\}$ be the eigenvalues of XX'. Then

$$\|\mathsf{K}_{\lambda_2}^{-1}\|\|\mathsf{K}_{\lambda_2}\| = \frac{\lambda_{\max}(\mathsf{K}_{\lambda_2})}{\lambda_{\min}(\mathsf{K}_{\lambda_2})} = \frac{n\lambda_2 + \max_{j\leq n} v_j}{n\lambda_2 + \min_{j\leq n} v_j}.$$

Thus a sufficient condition for $\|\mathsf{K}_{\lambda_2}^{-1}\|\|\mathsf{K}_{\lambda_2}\| \lesssim 1$ is that $\max_{j\leq n} v_j \lesssim n\lambda_2$, which is equivalent to $\|S\| \lesssim \lambda_2$ since $S = X'X/n$.

2.2. *Computation of* $\mathsf{K}_{\lambda_2}^{1/2}$. Let $X = MDV$ be the singular value decomposition of X: where D is an $n \times p$ rectangular diagonal matrix with square roots of the non-zero eigenvalues of XX' on the diagonal, and M is an $n \times n$ matrix whose columns are the eigenvectors of the $n \times n$ matrix XX'. Then from $M'M = I_n$ and $VV' = I_p$, $\mathsf{P}_{\lambda_2} = MD(D'D + n\lambda_2 I)^{-1}D'M'$. Hence $\mathsf{K}_{\lambda_2}^{1/2} = MAM'$, where $A = [I_n - D(D'D + n\lambda_2 I)^{-1}D']^{1/2}$. Note that $D(D'D + n\lambda_2 I)^{-1}D'$ is an $n \times n$ diagonal matrix with the jth diagonal entry as

$$\frac{v_j}{v_j + n\lambda_2},$$

where v_j is the jth largest eigenvalue of XX'. This implies that the jth diagonal entry of A is

$$a_j = \left(\frac{n\lambda_2}{v_j + n\lambda_2}\right)^{1/2}.$$

Department of Economics,
Massachusetts Institute of Technology,
Cambridge, MA 02139, USA.
E-mail: vchern@mit.edu

Booth School of Business,
University of Chicago,
Chicago, IL 60637, USA.
E-mail: Christian.Hansen@chicagobooth.edu

Department of Mathematics,
University of Maryland.
College Park, MD 20742, USA.
E-mail: yuanliao@umd.edu

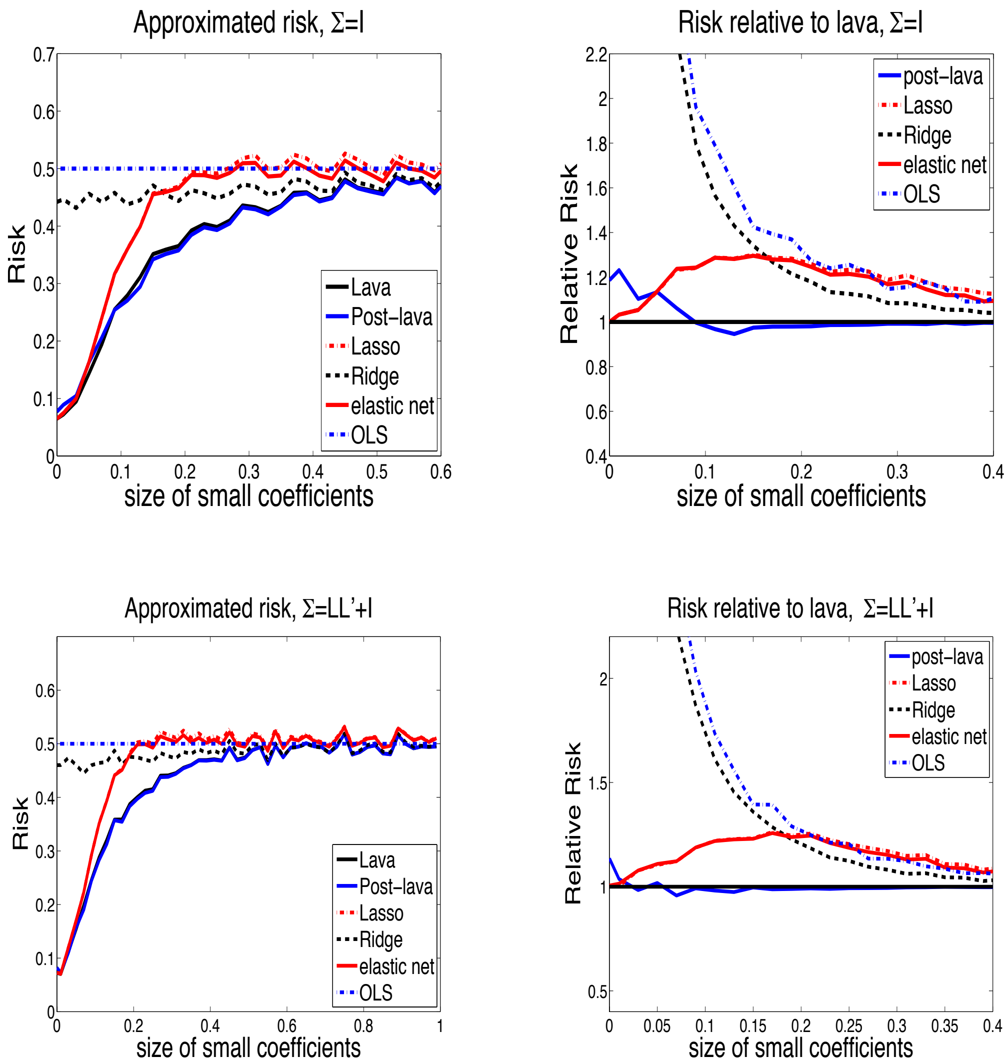

FIG 1. *Risk comparison with tuning done by minimizing SURE in the fixed design simulations with $p = n/2$. In this figure, we report simulation estimates of risk functions of ols, lava, post-lava, ridge, lasso, and elastic net in a Gaussian regression model with "sparse+dense" signal structure over the regression coefficients. We select tuning parameters for each method by minimizing SURE. The size of "small coefficients" (q) is shown on the horizontal axis. The size of these coefficients directly corresponds to the size of the "dense part" of the signal, with zero corresponding to the exactly sparse case. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_e)/\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{lava}})$.

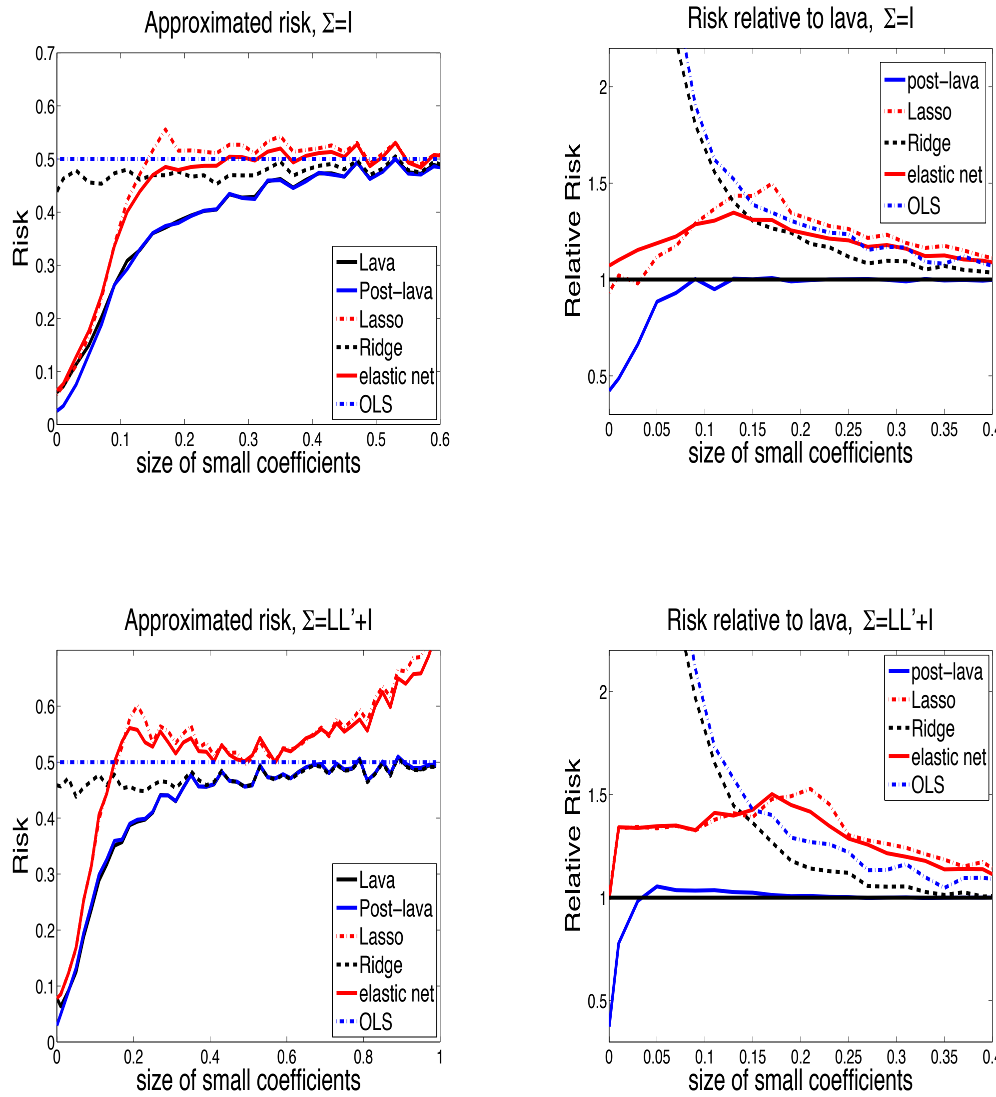

FIG 2. *Risk comparison with tuning done by 5-fold cross-validation in the fixed design simulations with $p = n/2$. We select tuning parameters for each method by 5-fold cross-validation. The size of "small coefficients" (q) is shown on the horizontal axis. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_e)/\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{lava}})$.

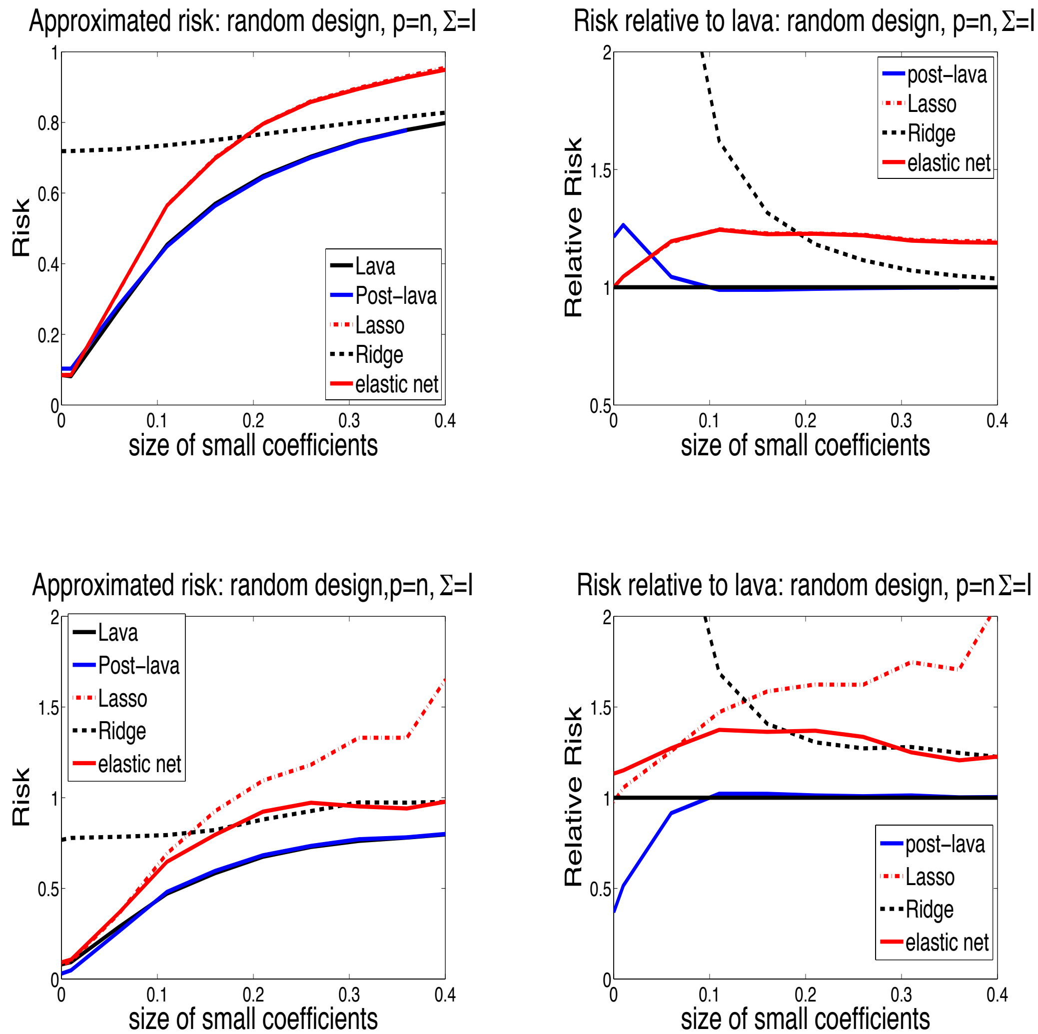

FIG 3. *Risk comparison in the random design simulations with $p = n$. In the upper panels, we report results with tuning parameters for each method selected by minimizing SURE. In the lower panels, we report results with tuning parameters for each method selected by 5-fold cross-validation. The size of "small coefficients" (q) is shown on the horizontal axis. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_e)/\mathrm{R}(\theta, \widehat{\theta}_{\text{lava}})$.

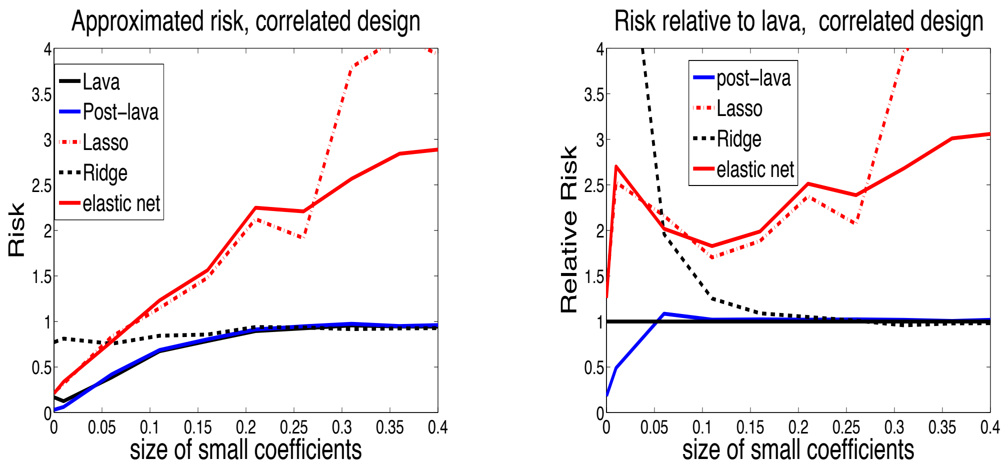

FIG 4. *Risk comparison in the highly correlated fixed design simulations with $p = 2n$. Here the first ten regressors are highly correlated with a common mutual correlation 0.89. We select tuning parameters for each method by 5-fold cross-validation. The size of "small coefficients" (q) is shown on the horizontal axis. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_e)/\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{lava}})$.

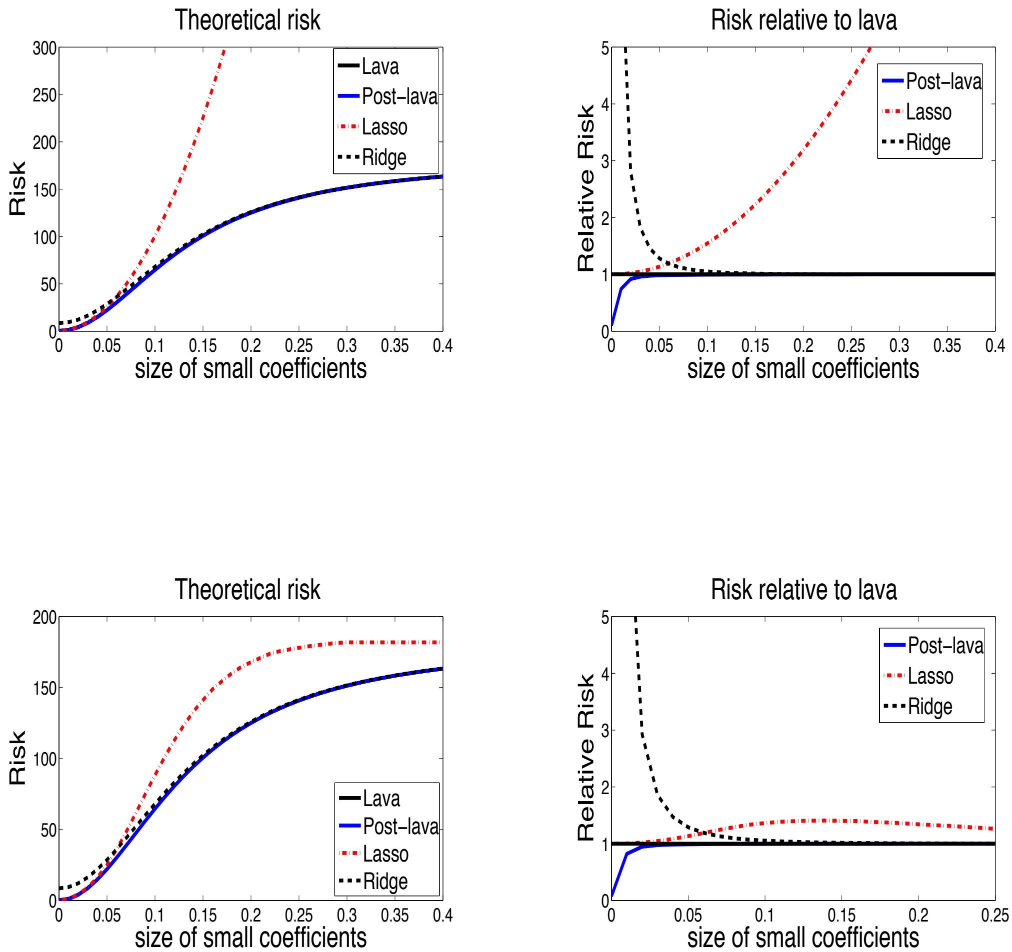

FIG 5. *Exact risk functions of lava, post-lava, lasso and ridge in the Gaussian sequence model with "sparse+dense" signal structure and* $p = 10,000$*. In the upper panels, we report results with tuning parameters for each method selected by the "canonical plug-in" choices given in the text. In the lower panels, we report results with the optimal (risk minimizing) choices of penalty levels. The size of "small coefficient" is shown on the horizontal axis. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_e)/\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{lava}})$*.*

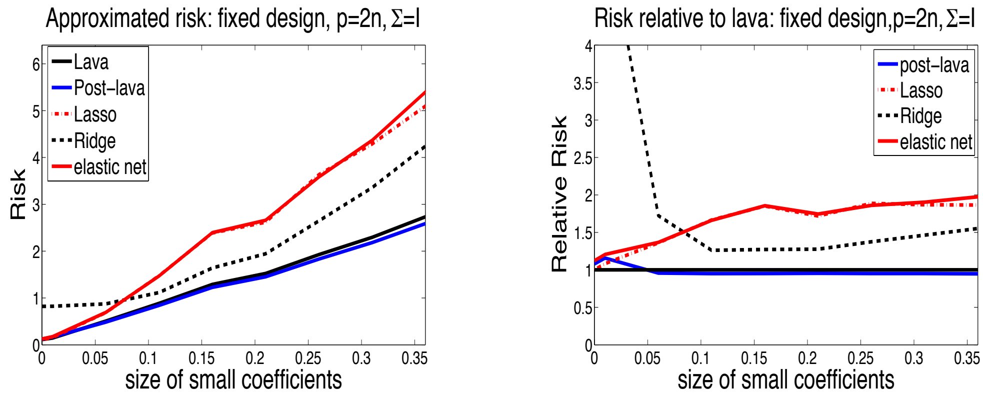

FIG 6. *Risk comparison in the random design simulations with $p = 2n$. Tuning parameters for each method are selected by minimizing a feasible version of SURE based on a conservative plug-in estimate of σ_u^2. The size of "small coefficients" (q) is shown on the horizontal axis. Relative risk plots the ratio of the risk of each estimator to the lava risk,* $\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{e}})/\mathrm{R}(\theta, \widehat{\theta}_{\mathrm{lava}})$.

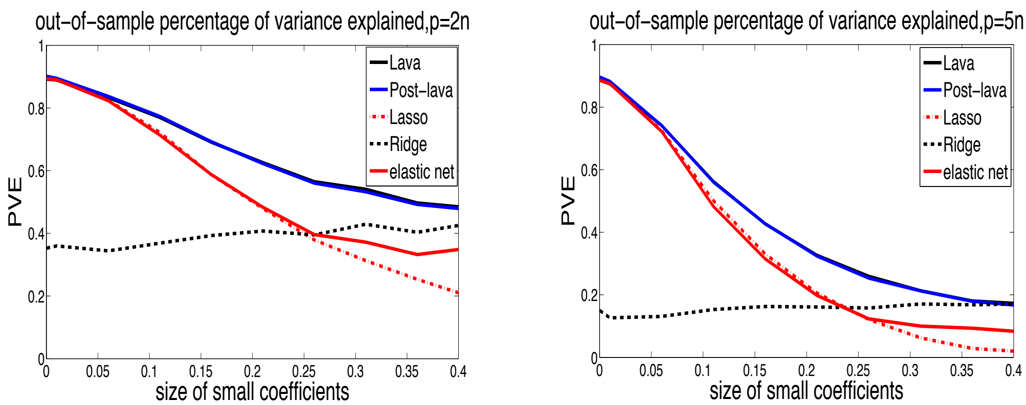

FIG 7. *Out-of-sample percentage of variation explained (PVE) in the random design simulation with $p = 2n$ and $p = 5n$. For each method, the tuning is chosen by minimizing 5-fold cross-validation.*