On the Validity of the Pairs Bootstrap for Lasso Estimators

Lorenzo Camponovo*
University of St.Gallen

(*Revise and Resubmit Biometrika)*

June 2014

Abstract

We study the validity of the pairs bootstrap for Lasso estimators in linear regression models with random covariates and heteroscedastic error terms. We show that the naive pairs bootstrap does not consistently estimate the distribution of the Lasso estimator. In particular, we identify two different sources for the failure of the bootstrap. First, in the bootstrap samples the Lasso estimator fails to correctly mimic the population moment condition satisfied by the regression parameter. Second, the bootstrap Lasso estimation criterion does not reproduce the sign of the zero coefficients with sufficient accuracy. To overcome these problems we introduce a modified pairs bootstrap procedure that consistently estimates the distribution of the Lasso estimator. Finally, we consider also the adaptive Lasso estimator. Also in this case, we show that the modified pairs bootstrap consistently estimates the distribution of the adaptive Lasso estimator. Monte Carlo simulations confirm a desirable accuracy of the modified pairs bootstrap procedure. These results show that when properly defined the pairs bootstrap may provide a valid approach for estimating the distribution of Lasso estimators.

AMS (2000) Subject Classification: Primary: 62J07; Secondary: 62G09, 62E20.
Keywords: Lasso Estimators, Pairs Bootstrap, Heteroscedastic Regression Models.

*E-mail: lorenzo.camponovo@unisg.ch. Address: Bodanstrasse 6, 9000 St.Gallen, Switzerland.

1 Introduction

1.1 Background and Results

Consider the linear regression model

$$y_t = \sum_{i=1}^{p} \beta_i x_{t,i} + \epsilon_t, \qquad t = 1, \dots n, \tag{1}$$

where y_t is the response variable, $x_t = (x_{t,1}, \dots, x_{t,p})'$ is a vector of covariates, ϵ_t is the disturbance of the regression, and $\beta = (\beta_1, \dots, \beta_p)'$ is the unknown parameter of interest. Throughout the paper, we assume that p is fixed, while n is large (the extension of our theoretical results to high-dimensional settings where the regression parameter is allowed to depend on the sample size n is briefly discussed in Remark 8 below). Furthermore, we also assume that the covariate vector x_t is random, and the error term ϵ_t may be related to x_t.

In the first part of the paper, we focus on the Lasso estimator $\hat{\beta}_n^L = (\hat{\beta}_{n,1}^L, \dots, \hat{\beta}_{n,p}^L)'$ of β defined as

$$\hat{\beta}_n^L = arg \min_u \sum_{t=1}^{n} (y_t - u'x_t)^2 + \lambda_n \sum_{i=1}^{p} |u_i|, \tag{2}$$

where $\lambda_n > 0$ is a tuning parameter. Since the introduction in Tibshirani (1996), Lasso estimators attained widespread applicability in different statistics and econometrics problems. Indeed, the Lasso approach enjoys many desirable properties. In particular, besides point estimation Lasso estimators may also perform valid model selection. Therefore, in several settings they may be preferred to alternative estimating procedures such as least squares estimators, among others. Important studies that investigate the model-consistency properties of the Lasso approach include Zhao and Yu (2006), Wainwright (2006), and Zou (2006); see also Bühlmann and van de Geer (2011) for a detailed discussion on the properties of Lasso estimators.

Knight and Fu (2000) derive the limit distribution of the Lasso estimator in linear regression models with nonrandom covariates and homoscedastic error terms. Wagener and Dette (2012) extend the results in Knight and Fu (2000) to linear regression models with nonrandom covariates and heteroscedastic error terms. By adapting their approach to our setting, we derive the limit distribution of the Lasso estimator in linear regression models with random covariates and heteroscedastic error terms. In particular, we show that under some regularity conditions the law of $T_n^L = \sqrt{n}(\hat{\beta}_n^L - \beta)$ converges weakly to $T^L = arg \min_{u \in \mathbb{R}^p} R^L(u)$, where $R^L(u)$ is a random process over $\mathbb{R}^p$. As highlighted in Chatterjee and Lahiri (2010), limit distributions of this form are quite complicated, and in practice it may be difficult to construct confidence sets or implement testing hypotheses on the regression parameter. Therefore, also in our setting it is suitable to verify whether we may approximate the sampling distribution of T_n^L with alternative methods. To this end, we investigate the validity of bootstrap approximations.

In this setting, the standard approach to bootstrapping is the pairs bootstrap; see, e.g., Freedman (1981). More precisely, let $(z_1, \ldots, z_n)$ be the observation sample, where for $t = 1, \ldots, n$, $z_t = (y_t, x_t')'$. The pairs bootstrap constructs random samples $(z_1^*, \ldots, z_n^*)$ by selecting from $(z_1, \ldots, z_n)$ with replacement. Let $\hat{\beta}_n^{L*}$ be the solution of (2) based on the bootstrap sample $(z_1^*, \ldots, z_n^*)$. The pairs bootstrap approximates the sampling distribution of T_n^L with the (conditional) distribution of $T_n^{L*} = \sqrt{n}(\hat{\beta}_n^{L*} - \hat{\beta}_n^L)$ given the observations $(z_1, \ldots, z_n)$. In this paper, we show that this approach does not consistently estimate the distribution of T_n^L. In particular, we identify two different sources for the failure of the pairs bootstrap. First, note that (typically) the regression parameter β satisfies the population moment condition $E[(y_t - x_t'\beta)x_t] = 0$. On the other hand, $E^*[(y_t^* - x_t'^*\hat{\beta}_n^L)x_t^*] \neq 0$, where E^* denote the expectation with respect to the distribution of the bootstrap sample conditional on the original sample. Therefore, the bootstrap moment condition based on the Lasso estimator does not correctly mimic the population moment condition. This distortion may heavily deteriorate the accuracy of the pairs bootstrap approximation. Finally, the second source of the failure of the pairs bootstrap is related to the incapability of the Lasso estimator of capturing the sign of the zero coefficients with sufficient high probability. Indeed, because of this inaccuracy the bootstrap Lasso estimation criterion does not properly reproduce the sign of the zero coefficients.

To overcome these problems, we introduce a modified pairs bootstrap procedure. In particular, first we propose to recenter the bootstrap statistic with respect to the least squares estimator $\hat{\beta}_n^{LS}$ instead of the Lasso estimator $\hat{\beta}_n^L$. Note that the least squares estimator $\hat{\beta}_n^{LS}$ is defined as the solution of (2) with $\lambda_n = 0$. Furthermore, it is important to highlight that $E^*[(y_t^* - x_t'^*\hat{\beta}_n^{LS})x_t^*] = 0$. Therefore, the least squares estimator correctly mimics the population moment condition $E[(y_t - x_t'\beta)x_t] = 0$. Second, we replace the standard Lasso estimation criterion with an adjusted bootstrap Lasso estimation criterion that properly reproduces the sign of the zero coefficients. By adopting these two corrections, we show that the modified pairs bootstrap consistently estimates the distribution of T_n^L.

Finally, in the last part of the paper we focus also on the adaptive Lasso estimator $\hat{\beta}_n^{AL} = (\hat{\beta}_{n,1}^{AL}, \ldots, \hat{\beta}_{n,p}^{AL})'$ of β defined as

$$\hat{\beta}_n^{AL} = arg \min_u \sum_{t=1}^{n} (y_t - u'x_t)^2 + \lambda_n \sum_{i=1}^{p} \lambda_{n,i} |u_i|, \tag{3}$$

where $\lambda_n > 0$ is a tuning parameter, $\lambda_{n,i} = 1/|\hat{\beta}_{n,i}|^\gamma$, $\gamma > 0$, and $\hat{\beta}_{n,i}$ is a root-n consistent estimator of β_i. Zou (2006) shows that in linear regression models with nonrandom covariates and homoscedastic error terms the adaptive Lasso possesses the so-called oracle properties. More precisely, the adaptive Lasso performs correct model selection. Furthermore, the adaptive Lasso estimates the nonzero coefficients of the regression parameter with the same efficiency (asymptotically) of the least squares estimator. Wagener and Dette (2012) extend the results in Zou (2006) to linear regression models with nonrandom covariates and heteroscedastic error terms. By adapting their approach to our setting, we show that the adaptive Lasso possesses the oracle properties also in linear regression models with random covariates and heteroscedastic error terms. Furthermore, we study the validity of the pairs bootstrap in approximating the sampling distribution of $T_n^{AL} = \sqrt{n}(\hat{\beta}_n^{AL} - \beta)$. In particular, we show that also in this case the modified pairs bootstrap consistently estimates the distribution of the adaptive Lasso estimator.

1.2 Contributions to the Literature

Existing literature mainly focuses on the validity of residual bootstrap methods for Lasso estimators in homoscedastic linear regression models, when the dimension p of the regression parameter β is fixed, and the sample size n is large. In particular, Chatterjee and Lahiri (2010) show that the naive residual bootstrap does not provide a valid method for approximating the sampling distribution of the Lasso estimator. On the other hand, Chatterjee and Lahiri (2011) prove that the naive residual bootstrap consistently estimates the distribution of the adaptive Lasso estimator. Moreover, Chatterjee and Lahiri (2011) also introduce a modified residual bootstrap procedure for the Lasso estimator. Finally, Hall, Lee and Park (2009) define an m-out-of-n residual bootstrap approach for the optimal selection of tuning parameters in adaptive Lasso settings. In this paper, we focus instead on the validity of the pairs bootstrap for Lasso estimators in more general heteroscedastic linear regression models. The more detailed contributions to the literature are as follows.

First, we show that the naive pairs bootstrap may have some issues in approximating the sampling distribution of the Lasso estimator. This result completes and extends the findings in Knight and Fu (2000) and Chatterjee and Lahiri (2010) on the inconsistency of the residual bootstrap for the Lasso estimator. In particular, Chatterjee and Lahiri (2010) show that the (only) source of the inconsistency of the residual bootstrap is related to the incapability of the Lasso estimator of capturing the sign of the zero coefficients with sufficient high probability. On the other hand, in this paper we identify two different sources for the failure of the pairs bootstrap: (i) the bootstrap moment condition based on the Lasso estimator does not correctly mimic the population moment condition satisfied by the regression parameter, and (ii) the pairs bootstrap Lasso estimation criterion does not properly reproduce the sign of the zero coefficients.

Second, to overcome these problems we introduce a modified pairs bootstrap procedure that consistently estimates the distribution of the Lasso estimator. The definition of this bootstrap method partly relies on the thresholding approach introduced in Chatterjee and Lahiri (2011). However, the implementation of the modified pairs bootstrap differs considerably from the residual bootstrap defined in Chatterjee and Lahiri (2011). In particular, to ensure the consistency of the bootstrap we propose to recenter the bootstrap statistic, and to adjust the bootstrap Lasso estimation criterion (see Equation (5) below for more details). To the best of our knowledge, this is the first paper in the Lasso literature that proposes adjustments of the bootstrap Lasso estimation criterion.

Third, we show that the modified pairs bootstrap can also be implemented to approximate the sampling distribution of the adaptive Lasso estimator. Since the adaptive Lasso estimator fails to correctly mimic the population moment condition satisfied by the regression parameter, the naive pairs bootstrap may have some issues in approximating also the sampling distribution of T_n^{AL}. To overcome this problem, we introduce the modified pairs bootstrap and prove its consistency. Note that also in this setting the modified pairs bootstrap differs considerably from the residual bootstrap method proposed in Chatterjee and Lahiri (2011), and the wild bootstrap approach introduced in Minnier, Tian and Cai (2011). In conclusion, these results show that when properly defined the pairs bootstrap may provide a valid approach for estimating the distribution of Lasso estimators.

1.3 Outline of the Paper

In Section 2, we consider the pairs bootstrap for the Lasso estimator. In particular, in Section 2.1 we consider the naive pairs bootstrap, while in Section 2.2 we introduce the modified pairs bootstrap. In Section 3, we focus on the pairs bootstrap for the adaptive Lasso estimator. In Section 4, we study through Monte Carlo simulations the accuracy of bootstrap methods. Finally, in Section 5 we provide some concluding remarks.

2 Pairs Bootstrap for the Lasso Estimator

2.1 The Naive Pairs Bootstrap

Consider the regression model (1) and the Lasso estimator $\hat{\beta}_n^L$ defined in (2). In this section, we study the validity of bootstrap methods in approximating the sampling distribution of $T_n^L = \sqrt{n}(\hat{\beta}_n^L - \beta)$. Before presenting the main results, we introduce the following assumption.

Assumption 2.1. *The vectors x_t are independent, with common distribution. Furthermore, let $E[\|z_t\|^4] < \infty$, where $\|\cdot\|$ is the Euclidean norm. The parameter β minimizes $E[(y_t - x_t'\beta)^2]$. Finally, let $C = E[x_t x_t']$ be positive definite, and the law of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t \epsilon_t$ converges weakly to normal with mean 0 and variance-covariance Ω.*

Assumption 2.1 is a set of mild conditions which are also required to prove the consistency of the pairs bootstrap in estimating the distribution of the least squares estimator in linear regression models with random covariates and heteroscedastic error terms; see, e.g., Freedman (1981). Note that Assumption 2.1 also implies that β satisfies the moment condition $E[(y_t - x_t'\beta)x_t] = 0$. In the next lemma we derive the limit distribution of T_n^L.

Lemma 2.1. *Let Assumption 2.1 hold. If $\frac{\lambda_n}{\sqrt{n}} \to \lambda_0 \geq 0$, then the law of $\sqrt{n}(\hat{\beta}_n^L - \beta)$ converges weakly to $T^L = arg\min_{u\in\mathbb{R}^p} R^L(u)$, where*

$$R^L(u) = -2u'W + u'Cu + \lambda_0 \sum_{i=1}^{p} [u_i sgn(\beta_i)\mathbb{I}(\beta_i \neq 0) + |u_i|\mathbb{I}(\beta_i = 0)],$$

$W \sim N(0, \Omega)$, and $\mathbb{I}(\cdot)$ is the indicator function.

Proof. To prove Lemma 2.1 we adopt the same arguments of the proof of Theorem 2 in Knight and Fu (2000). More precisely, first we show that T_n^L minimizes a particular objective function $R_n^L(u)$. Then, we compute the limit of $R_n^L(u)$ denoted by $R^L(u)$. Finally, using the results in Geyer (1994), we conclude that the law of T_n^L converges weakly to $T^L = arg\min_{u\in\mathbb{R}^p} R^L(u)$.

To this end, let

$$R_n^L(u) = \sum_{t=1}^{n} \left[(\epsilon_t - u'x_t/\sqrt{n})^2 - \epsilon_t^2\right] + \lambda_n \sum_{i=1}^{p} \left[|\beta_i + u_i/\sqrt{n}| - |\beta_i|\right].$$

Note that $R_n^L(u)$ is minimized at $\sqrt{n}(\hat{\beta}_n^L - \beta)$. Consider the first term of $R_n^L(u)$. In particular, we have

$$\sum_{t=1}^{n} \left[(\epsilon_t - u'x_t/\sqrt{n})^2 - (\epsilon_t)^2\right] = -\frac{2}{\sqrt{n}} \sum_{t=1}^{n} u'x_t\epsilon_t + \frac{1}{n}\sum_{t=1}^{n} u'x_t x_t' u.$$

Note that $\frac{1}{n}\sum_{t=1}^{n} x_t x_t'$ converges in probability to C. Furthermore, the law of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\epsilon_t$ converges weakly to normal with mean 0 and variance-covariance Ω.

Next, consider the second term of $R_n^L(u)$. Then, as in Knight and Fu (2000) we can show that the limit of this term is

$$\lambda_0 \sum_{i=1}^{p} [u_i \mathrm{sgn}(\beta_i)\mathbb{I}(\beta_i \neq 0) + |u_i|\mathbb{I}(\beta_i = 0)].$$

Therefore, by Geyer (1994) Theorem 2.1 is established. □

By comparing the results in our Lemma 2.1 with those in Theorem 2 in Knight and Fu (2000), and Lemma 3.1 in Wagener and Dette (2012), we can observe that there are some obvious similarities between the limit distribution of the Lasso estimator in our setting and in linear regression models with nonrandom covariates. Furthermore, when $\lambda_0 = 0$, then $arg\min_{u\in\mathbb{R}^p} R(u) = C^{-1}W \sim N(0, C^{-1}\Omega C^{-1})$. Therefore, in this case we have the same limit distribution of the least squares estimator; see, e.g., Freedman (1981). On the other hand, when $\lambda_0 > 0$, then it may be quite complicated to construct confidence sets or implement testing hypotheses on the regression parameter β using the limit distribution of T_n^L.

After presenting the limit distribution of T_n^L, we consider bootstrap approximations. In this setting, the standard approach to bootstrapping is the pairs bootstrap. More precisely, let $(z_1^*, \ldots, z_n^*)$ be a random sample selected from $(z_1, \ldots, z_n)$ with replacement. We introduce the bootstrap Lasso estimator $\hat{\beta}_n^{L*}$ defined as

$$\hat{\beta}_n^{L*} = arg\min_u \sum_{t=1}^n (y_t^* - u'x_t^*)^2 + \lambda_n \sum_{i=1}^p |u_i|.$$

Finally, we approximate the sampling distribution of $T_n^L = \sqrt{n}(\hat{\beta}_n^L - \beta)$ with the (conditional) distribution of $T_n^{L*} = \sqrt{n}(\hat{\beta}_n^{L*} - \hat{\beta}_n^L)$ given the observations $(z_1, \ldots, z_n)$.

To verify the validity of the pairs bootstrap approximation, we apply the same approach adopted in the proof of Theorem 2.1, and in Section 4 in Knight and Fu (2000). More precisely, first we show that T_n^{L*} minimizes a particular objective function $R_n^{L*}(u)$. Then, we compute the limit of $R_n^{L*}(u)$ denoted by $R^{L*}(u)$. Finally, we compare $R^{L*}(u)$ with $R^L(u)$.

To this end, let

$$R_n^{L*}(u) = \sum_{t=1}^n \left[(\hat{\epsilon}_t^{L*} - u'x_t^*/\sqrt{n})^2 - (\hat{\epsilon}_t^{L*})^2\right] + \lambda_n \sum_{i=1}^p \left[|\hat{\beta}_{n,i}^L + u_i/\sqrt{n}| - |\hat{\beta}_{n,i}^L|\right],$$

where $\hat{\epsilon}_t^{L*} = y_t^* - x_t'^*\hat{\beta}_n^L$. Note that $R_n^{L*}(u)$ is minimized at $\sqrt{n}(\hat{\beta}_n^{L*} - \hat{\beta}_n^L)$. Consider the first term of $R_n^{L*}(u)$. In particular, we have

$$\begin{aligned}
\sum_{t=1}^n \left[(\hat{\epsilon}_t^{L*} - u'x_t^*/\sqrt{n})^2 - (\hat{\epsilon}_t^{L*})^2\right] &= -\frac{2}{\sqrt{n}}\sum_{t=1}^n u'x_t^*\hat{\epsilon}_t^{L*} + \frac{1}{n}\sum_{t=1}^n u'x_t^*x_t'^*u \\
&= -\frac{2}{\sqrt{n}}\sum_{t=1}^n u'x_t^*(\hat{\epsilon}_t^{L*} - \hat{\epsilon}_t^{LS*}) - \frac{2}{\sqrt{n}}\sum_{t=1}^n u'x_t^*\hat{\epsilon}_t^{LS*} + \frac{1}{n}\sum_{t=1}^n u'x_t^*x_t'^*u \\
&= -\frac{2}{\sqrt{n}}\sum_{t=1}^n u'x_t^*x_t'^*(\hat{\beta}_n^{LS} - \hat{\beta}_n^L) - \frac{2}{\sqrt{n}}\sum_{t=1}^n u'x_t^*\hat{\epsilon}_t^{LS*} + \frac{1}{n}\sum_{t=1}^n u'x_t^*x_t'^*u,
\end{aligned}$$

where $\hat{\epsilon}_t^{LS*} = y_t^* - x_t'^*\hat{\beta}_n^{LS}$. Note that under Assumption 2.1, we can show that $\frac{1}{n}\sum_{t=1}^n x_t^*x_t'^*$ converges in conditional probability to C, while the conditional law of $\frac{1}{\sqrt{n}}\sum_{t=1}^n x_t^*\hat{\epsilon}_t^{LS*}$ converges weakly to normal with mean 0 and variance-covariance Ω; see, e.g., Freedman (1981).

Suppose that $\sqrt{n}(\hat{\beta}_n^L - \beta)$ converges weakly to $U^L = (U_1^L, \ldots, U_p^L)'$. Then, using the same arguments in Section 4 in Knight and Fu (2000) we can show that the second term of $R_n^{L*}(u)$ converges to

$$\lambda_0 \sum_{i=1}^{p} [u_i \mathrm{sgn}(\beta_i) \mathbb{I}(\beta_i \neq 0) + (|u_i + U_i^L| - |U_i^L|) \mathbb{I}(\beta_i = 0)].$$

Furthermore, suppose that $\sqrt{n}(\hat{\beta}_n^{LS} - \hat{\beta}_n^L)$ converges weakly to $U^{LS,L} = (U_1^{LS,L}, \ldots, U_p^{LS,L})'$. Then, we can conclude that $R_n^{L*}(u)$ converges to

$$R^{L*}(u) = -2u'CU^{LS,L} - 2u'W + u'Cu + \lambda_0 \sum_{i=1}^{p} [u_i \mathrm{sgn}(\beta_i) \mathbb{I}(\beta_i \neq 0) + (|u_i + U_i^L| - |U_i^L|) \mathbb{I}(\beta_i = 0)],$$

where $W \sim N(0, \Omega)$. By comparing $R^L(u)$ with $R^{L*}(u)$, we can note two important differences. First, the term $-2u'CU^{LS,L}$ appears only in $R^{L*}(u)$, while it is not included in the definition of $R^L(u)$. Second, the penalization of the zero coefficients in $R^L(u)$ and $R^{L*}(u)$ is slightly different. Note that the source of the term $-2u'CU^{LS,L}$ in the definition of $R^{L*}(u)$ is related to the distortion $E^*[(y_t^* - x_t'^* \hat{\beta}_n^L) x_t^*] \neq 0$. Furthermore, the different penalization of the zero coefficients in the definition of $R^{L*}(u)$ is instead related to the incapability of the Lasso estimator of capturing the sign of the zero coefficients with sufficient high probability. Because of these two differences the naive pairs bootstrap could provide a very poor approximation of the distribution of T_n^L.

This result completes and extends the findings in Knight and Fu (2000) and Chatterjee and Lahiri (2010) on the inconsistency of the residual bootstrap for the Lasso estimator. In particular, it is interesting to highlight that unlike Chatterjee and Lahiri (2010) we identify two different sources for the failure of the pairs bootstrap. In the next section, we show how to overcome these problems.

2.2 The Modified Pairs Bootstrap

In this section, we introduce a modified pairs bootstrap procedure that consistently estimates the distribution of $T_n^L = \sqrt{n}(\hat{\beta}_n^L - \beta)$. Our method partly relies on the thresholding approach introduced in Chatterjee and Lahiri (2011). Therefore, before presenting our procedure, first we briefly review the main idea of the modified residual bootstrap introduced in Chatterjee and Lahiri (2011).

In a linear regression model with nonrandom covariates and homoscedastic error terms the standard approach to bootstrapping is the residual bootstrap. Chatterjee and Lahiri (2010) show that the naive residual bootstrap does not consistently estimate the distribution of the Lasso estimator. In particular, they show that the residual bootstrap fails to reproduce the sign of the zero coefficients with sufficient accuracy in the formulation of the bootstrap Lasso estimation criterion. To overcome this problem, Chatterjee and Lahiri (2011) propose a thresholding procedure. More precisely, in the implementation of the residual bootstrap they propose to replace $\hat{\beta}_n^L$ with the modified Lasso estimator $\tilde{\beta}_n^L = (\tilde{\beta}_{n,1}^L, \ldots, \tilde{\beta}_{n,p}^L)'$, where

$$\tilde{\beta}_{n,i}^L = \hat{\beta}_{n,i}^L \cdot \mathbb{I}\left(|\hat{\beta}_{n,i}^L| \geq a_n\right),$$

and a_n denote a sequence of real numbers such that

$$a_n + (n^{-1/2} \log n) a_n^{-1} \to 0, \tag{4}$$

as $n \to \infty$. Examples of sequences that satisfy condition (4) include sequences of the form $a_n = cn^{-\delta}$, for $c \in (0, \infty)$ and $\delta \in (0, 1/2)$. It is important to highlight that the thresholding approach has no impact on the nonzero coefficients of β for n large. Indeed, if $\beta_i \neq 0$, then $|\hat{\beta}_{n,i}^L| > a_n$ for large n with high probability. On the other hand, this approach shrinks to 0 with high probability the zero coefficients of β. Indeed, if $\beta_i = 0$, then $|\hat{\beta}_{n,i}^L| < a_n$ for large n with high probability. Therefore, the modified Lasso estimator is able to capture the target sign of the zero coefficients with high probability. Chatterjee and Lahiri (2011) prove that the modified residual bootstrap consistently estimates the distribution of T_n^L.

Obviously, the thresholding approach cannot be applied to the pairs bootstrap. Indeed, the pairs bootstrap constructs the random samples in a fully nonparametric way. On the other hand, it is still possible to adjust the bootstrap Lasso estimation criterion in order to properly reproduce the sign of the zero coefficients. In particular, to introduce a valid pairs bootstrap procedure we propose the following corrections. First, to overcome the distortion $E^*[(y_t^* - x_t'^* \hat{\beta}_n^L) x_t^*] \neq 0$, we propose to recenter the bootstrap statistic with respect to $\hat{\beta}_n^{LS}$ instead of $\hat{\beta}_n^L$. Indeed, the least square estimator satisfies the bootstrap moment condition $E^*[(y_t^* - x_t'^* \hat{\beta}_n^{LS}) x_t^*] = 0$, which mimics the population moment condition $E[(y_t - \beta' x_t) x_t] = 0$. Second, given a pairs bootstrap sample $(z_1^*, \ldots, z_n^*)$, we introduce the modified Lasso estimator $\tilde{\beta}_n^{L*}$ defined as

$$\tilde{\beta}_n^{L*} = arg \min_u \sum_{t=1}^n (y_t^* - u' x_t^*)^2 + \lambda_n \sum_{i=1}^p |u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n)|, \tag{5}$$

where the sequence a_n satisfies condition (4). Note that in the penalization term in (5) we recenter with respect to $\hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n)$. The recentering has no impact on the nonzero coefficients of β for n large. Indeed, if $\beta_i \neq 0$, then $|\hat{\beta}_{n,i}^{LS}| > a_n$ for large n with high probability, and consequently $u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n) = u_i$. On the other hand, if $\beta_i = 0$, then $|\hat{\beta}_{n,i}^{LS}| \leq a_n$ for large n with high probability, and consequently $u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n) = u_i - \hat{\beta}_{n,i}^{LS}$. Therefore, for large n the penalization term in (5) recenters only the zero coefficients of β. Finally, we propose to approximate the sampling distribution of T_n^L with the (conditional) bootstrap distribution of

$\tilde{T}_n^{L*} = \sqrt{n}(\tilde{\beta}_n^{L*} - \hat{\beta}_n^{LS})$ given the observations $(z_1, \ldots, z_n)$. In the next theorem, we prove the validity of the modified pairs bootstrap.

Theorem 2.1. *Let Assumption 2.1 hold. If* $\frac{\lambda_n}{\sqrt{n}} \to \lambda_0 \geq 0$*, then the conditional law of* $\sqrt{n}(\tilde{\beta}_n^{L*} - \hat{\beta}_n^{LS})$ *converges weakly to* $T^L = arg\min_{u \in \mathbb{R}^p} R^L(u)$*, where*

$$R^L(u) = -2u'W + u'Cu + \lambda_0 \sum_{i=1}^{p} [u_i sgn(\beta_i)\mathbb{I}(\beta_i \neq 0) + |u_i|\mathbb{I}(\beta_i = 0)],$$

and $W \sim N(0, \Omega)$.

Proof. We adopt the same arguments in the proof of our Theorem 2.1, and Theorem 3.1 in Chatterjee and Lahiri (2011). More precisely, let

$$\tilde{R}_n^{L*}(u) = \sum_{t=1}^{n} \left[(\hat{\epsilon}_t^{LS*} - u'x_t^*/\sqrt{n})^2 - (\hat{\epsilon}_t^{LS*})^2 \right] + \lambda_n \sum_{i=1}^{p} \left[|\tilde{\beta}_{n,i}^{LS} + u_i/\sqrt{n}| - |\tilde{\beta}_{n,i}^{LS}| \right],$$

where $\tilde{\beta}_{n,i}^{LS} = \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| > a_n)$. Note that $\tilde{R}_n^{L*}(u)$ is minimized at $\sqrt{n}(\tilde{\beta}_n^{L*} - \hat{\beta}_n^{LS})$. Consider the first term of $\tilde{R}_n^{L*}(u)$. In particular, we have

$$\sum_{t=1}^{n} \left[(\hat{\epsilon}_t^{LS*} - u'x_t^*/\sqrt{n})^2 - (\hat{\epsilon}_t^{LS*})^2 \right] = -\frac{2}{\sqrt{n}} \sum_{t=1}^{n} u'x_t^* \hat{\epsilon}_t^{LS*} + \frac{1}{n} \sum_{t=1}^{n} u'x_t^* x_t'^* u.$$

By Theorem 3.1 in Freedman (1981), $\frac{1}{n}\sum_{t=1}^{n} x_t^* x_t'^*$ converges in conditional probability to C. Furthermore, also by Theorem 3.1 in Freedman (1981) the conditional law of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t^* \hat{\epsilon}_t^{LS*}$ converges weakly to normal with mean 0 and variance-covariance Ω.

Next, consider the second term in $\tilde{R}_n^{L*}(u)$. Note that as highlighted in the proof of Theorem 3.1 in Chatterjee and Lahiri (2011), for the nonzero coefficients $\beta_i \neq 0$, we have $\text{sgn}(\tilde{\beta}_{n,i}^{LS}) = \text{sgn}(\beta_i)$ for large n. Furthermore, for the zero coefficients $\beta_i = 0$, we have $\tilde{\beta}_{n,i}^{LS} = 0$ for large n. Therefore, the second term of $\tilde{R}_n^{L*}(u)$ converges to

$$\lambda_0 \sum_{i=1}^{p} [u_i \text{sgn}(\beta_i)\mathbb{I}(\beta_i \neq 0) + |u_i|\mathbb{I}(\beta_i = 0)].$$

Thus, we can conclude that $\tilde{R}_n^{L*}(u)$ converges weakly to

$$R^L(u) = -2u'W + u'Cu + \lambda_0 \sum_{i=1}^{p} [u_i \text{sgn}(\beta_i)\mathbb{I}(\beta_i \neq 0) + |u_i|\mathbb{I}(\beta_i = 0)],$$

where $W \sim N(0, \Omega)$. Using the results in Geyer (1994) Theorem 2.1 is established. □

Theorem 2.1 shows that the modified pairs bootstrap provides a valid approach for approximating the sampling distribution of $T_n^L = \sqrt{n}(\hat{\beta}_n^L - \beta)$. Furthermore, by adapting the results in Corollary 3.2 and Theorem 3.3 in Chatterjee and Lahiri (2011) to our setting, we can show that the modified pairs bootstrap can also be applied for the implementation of testing hypotheses on the regression parameter, and the variance estimation of the Lasso estimator.

Remark 1. The proposed modified pairs bootstrap differs considerably from the residual bootstrap procedure introduced in Chatterjee and Lahiri (2011). In particular, in (5) we propose to recenter the bootstrap Lasso estimation criterion with respect to $\hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n)$. This recentering provides the key correction for ensuring the consistency of the pairs bootstrap. To the best of our knowledge, this is the first paper in the Lasso literature that proposes adjustments of the bootstrap Lasso estimation criterion. Finally, note that the optimization problem in (5) can be solved with conventional algorithms adopted for the (standard) Lasso estimator. Indeed, using the substitution $\gamma_i = u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n)$, we obtain the (standard) Lasso estimation criterion.

Remark 2. In this remark, we better clarify the corrections introduced in the definition of the modified pairs bootstrap. First, to overcome the distortion $E^*[(y_t^* - x_t'^*\hat{\beta}_n^L)x_t^*] \neq 0$, we recenter the bootstrap statistic with respect to the least squares estimator $\hat{\beta}_n^{LS}$. Note that the least squares estimator satisfies $E^*[(y_t^* - x_t'^*\hat{\beta}_n^{LS})x_t^*] = 0$, and corresponds to the "true" value for the random bootstrap samples. Finally, because of this recentering we adjust the bootstrap Lasso estimation criterion using the term $\hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(|\hat{\beta}_{n,i}^{LS}| \leq a_n)$. Indeed, with this adjustment the bootstrap Lasso estimation criterion shrinks the zero coefficients of β to $\hat{\beta}_{n,i}^{LS}$ (the "true" value for the bootstrap samples). This adjustment exactly mimics the (standard) Lasso estimation criterion that shrinks the zero coefficients of β to 0 (the "true" value for the original sample).

Remark 3. In Theorem 2.1, we prove the validity of the modified pairs bootstrap for the Lasso estimator in linear regression models with random covariates and heteroscedastic error terms. Obviously, this result applies also to linear regression models with nonrandom covariates and homoscedastic error terms. With some appropriate modifications, we could also extend our finding to generalized linear models. However, in this paper we do not pursue this direction.

Remark 4. The accuracy of the Lasso estimator may heavily depend on the selection of λ_n. Furthermore, in the definition of the modified pairs bootstrap we have also to select the sequence a_n. Using the results in Theorem 2.1, we can introduce a data-driven method for the selection of these tuning parameters in the spirit of Hall, Lee and Park (2009), and Chatterjee and Lahiri (2011). The key rationale of our procedure is to select the optimal tuning parameters that minimize the estimated mean squared error of the Lasso estimator $\hat{\beta}_n^L$. We briefly describe the data-driven approach. For $\lambda_n \in [0, +\infty)$ and $a_n \in (0, +\infty)$, let $\tilde{\beta}_n^{L*} = \tilde{\beta}_n^{L*}(\lambda_n, a_n)$ denote the bootstrap modified Lasso estimator. Using Theorem 2.1, we can estimate the mean squared error $E[\|\hat{\beta}_n^L - \beta\|^2]$ of $\hat{\beta}_n^L$ by

$$\phi(\lambda_n, a_n) = E^*[\|\tilde{\beta}_n^L(\lambda_n, a_n) - \hat{\beta}_n^{LS}\|^2]. \tag{6}$$

Finally, we select the optimal values $(\hat{\lambda}_n, \hat{a}_n)$ that minimize the estimated mean squared error (6). In Section 4 below, we study through Monte Carlo simulations the accuracy of the modified pairs bootstrap combined with this data-driven method for the selection of the tuning parameters.

3 Pairs Bootstrap for the Adaptive Lasso Estimator

Consider the adaptive Lasso estimator $\hat{\beta}_n^{AL}$ defined in (3). To simplify the presentation of our results, we consider only the weights $\lambda_{n,i} = 1/|\hat{\beta}_{n,i}^{LS}|^\gamma$, with $\gamma = 1$. However, with some slight modifications the results in Theorems 3.1 and 3.1 below can also be extended to the more general penalization introduced in (3). In this section, we analyze the validity of the pairs bootstrap in approximating the sampling distribution of $T_n^{AL} = \sqrt{n}(\hat{\beta}_n^{AL} - \beta)$.

Before presenting the main results, first we introduce some notation. Let $A = \{i : \beta_i \neq 0\}$, and let $A^c = \{i : \beta_i = 0\}$. Let $\beta_A = (\beta_{A,1}, \ldots, \beta_{A,q})'$ and $\beta_{A^c} = (\beta_{A^c,1}, \ldots, \beta_{A^c,p-q})'$ denote the sub-vectors of the nonzero and zero coefficients of β, respectively, where $q \leq p$. Similarly, let $\hat{\beta}_{n,A}^{LS}$ and $\hat{\beta}_{n,A}^{AL}$ denote the least squares and the adaptive Lasso estimators of β_A, respectively. Furthermore, let $\hat{\beta}_{n,A^c}^{LS}$ and $\hat{\beta}_{n,A^c}^{AL}$ denote the least squares and the adaptive Lasso estimators of β_{A^c}, respectively. Finally, let V_A be the asymptotic variance-covariance matrix of $\hat{\beta}_{n,A}^{LS}$, and let $\hat{A}_n^{AL} = \{i : \hat{\beta}_{n,i}^{AL} \neq 0\}$. In the next lemma, we present the oracle properties of the adaptive Lasso for heteroscedastic linear regression models with random covariates.

Lemma 3.1. *Let Assumption 2.1 hold. If $\lambda_n \to +\infty$ and $\frac{\lambda_n}{\sqrt{n}} \to 0$, then:*

(a) Variable Selection: $\lim_{n\to\infty} P(\hat{A}_n^{AL} = A) = 1$.

(b) Limit Distribution: The law of $\sqrt{n}\left(\hat{\beta}_{n,A}^{AL} - \beta_A\right)$ converges weakly to normal with mean 0 and variance-covariance V_A.

Proof. Lemma 3.1 follows directly from Theorem 1 in Audrino and Camponovo (2013). □

Zou (2006) shows that in linear regression models with nonrandom covariates and homoscedastic error terms the adaptive Lasso possesses the so-called oracle properties. More precisely, the adaptive Lasso performs correct model selection. Moreover, the adaptive Lasso estimates the nonzero coefficients of the regression parameter with the same efficiency (asymptotically) of the least squares estimator. Lemma 3.1 shows that the adaptive Lasso satisfies these two properties also in linear regression models with random covariates and heteroscedastic error terms.

After presenting the oracle properties, we consider bootstrap approximations. Let $(z_1^*, \ldots, z_n^*)$ be a random bootstrap sample selected from $(z_1, \ldots, z_n)$ with replacement. Note that also for the adaptive Lasso estimator $E^*[(y_t^* - x_t'^* \hat{\beta}_n^{AL})x_t^*] \neq 0$. Therefore, to overcome this distortion also in this setting we propose to recenter the adaptive Lasso bootstrap statistic with respect to $\hat{\beta}_n^{LS}$ instead of $\hat{\beta}_n^{AL}$. Unlike the least squares and the Lasso estimators, the adaptive Lasso captures the correct sign of the zero coefficients with high probability. However, since we recenter the bootstrap statistic with respect to $\hat{\beta}_n^{LS}$ instead of $\hat{\beta}_n^{AL}$, also in this setting we have to modify the bootstrap adaptive Lasso estimation criterion. More precisely, we introduce the modified bootstrap adaptive Lasso estimator $\tilde{\beta}_n^{AL*}$ defined as

$$\tilde{\beta}_n^{AL*} = arg \min_u \sum_{t=1}^{n} (y_t^* - u' x_t^*)^2 + \lambda_n \sum_{i=1}^{p} \lambda_{n,i}^* |u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0)|, \tag{7}$$

where $\lambda_{n,i}^* = 1/|\hat{\beta}_{n,i}^{LS*}|$, and $\hat{\beta}_n^{LS*}$ is the bootstrap least squares estimator. Note that in the penalization term in (7) we recenter with respect to $\hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0)$. The recentering has no impact on the nonzero coefficients of β for n large. Indeed, if $\beta_i \neq 0$, then $|\hat{\beta}_{n,i}^{AL}| > 0$ for large n with high probability, and consequently $u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0) = u_i$. On the other hand, if $\beta_i = 0$, then $\hat{\beta}_{n,i}^{AL} = 0$ for large n with high probability, and consequently $u_i - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0) = u_i - \hat{\beta}_{n,i}^{LS}$. Therefore, the penalization term in (7) recenters only the zero coefficients of β. Finally, we approximate the sampling distribution of T_n^{AL} with the (conditional) bootstrap distribution of $\tilde{T}_n^{AL*} = \sqrt{n}(\tilde{\beta}_n^{AL*} - \hat{\beta}_n^{LS})$ given the observations $(z_1, \ldots, z_n)$. Let $\tilde{\beta}_{n,A}^{AL*}$ and $\tilde{\beta}_{n,A^c}^{AL*}$ denote the modified bootstrap adaptive Lasso estimators of β_A and β_{A^c}, respectively. In the next theorem, we establish the validity of the modified pairs bootstrap in approximating the sampling distribution of the adaptive Lasso estimator.

Theorem 3.1. *Let Assumption 2.1 hold. If* $\lambda_n \to +\infty$ *and* $\frac{\lambda_n}{\sqrt{n}} \to 0$, *then:*

(a) Zero Coefficients: The conditional law of $\sqrt{n}\left(\tilde{\beta}_{n,A^c}^{AL*} - \hat{\beta}_{n,A^c}^{LS}\right)$ *converges in probability to* 0.

(b) Nonzero Coefficients: The conditional law of $\sqrt{n}\left(\tilde{\beta}_{n,A}^{AL*} - \hat{\beta}_{n,A}^{LS}\right)$ *converges weakly to normal with mean* 0 *and variance-covariance* V_A.

Proof. We adopt the same arguments in the proof of our Theorem 2.1, and Theorem 2 in Zou

(2006). More precisely, let

$$
\begin{aligned}
\tilde{R}_n^{AL*}(u) &= \sum_{t=1}^{n} \left[(\hat{\epsilon}_t^{LS*} - u' x_t^* / \sqrt{n})^2 - (\hat{\epsilon}_t^{LS*})^2 \right] \\
&+ \lambda_n \sum_{i=1}^{p} \lambda_{n,i}^* \left[|\hat{\beta}_{n,i}^{LS} - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0) + u_i/\sqrt{n}| - |\hat{\beta}_{n,i}^{LS} - \hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0)| \right].
\end{aligned}
$$

Note that $\tilde{R}_n^{AL*}(u)$ is minimized at $\sqrt{n}(\tilde{\beta}_n^{AL*} - \hat{\beta}_n^{LS})$. Consider the first term of $\tilde{R}_n^{AL*}(u)$. In particular, we have

$$
\sum_{t=1}^{n} \left[(\hat{\epsilon}_t^{LS*} - u' x_t^* / \sqrt{n})^2 - (\hat{\epsilon}_t^{LS*})^2 \right] = -\frac{2}{\sqrt{n}} \sum_{t=1}^{n} u' x_t^* \hat{\epsilon}_t^{LS*} + \frac{1}{n} \sum_{t=1}^{n} u' x_t^* x_t'^* u.
$$

By Theorem 3.1 in Freedman (1981), $\frac{1}{n}\sum_{t=1}^{n} x_t^* x_t'^*$ converges in conditional probability to C. Furthermore, also by Theorem 3.1 in Freedman (1981) the conditional law of $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t^* \hat{\epsilon}_t^*$ converges weakly to normal with mean 0 and variance-covariance Ω.

Consider the second term of $\tilde{R}_n^{AL*}(u)$. Note that by Theorem 3.1 in Freedman (1981), the conditional law of $\hat{\beta}_n^{LS*} - \hat{\beta}_n^{LS}$ converges in probability to 0. If $\beta_i \neq 0$, then $\hat{\beta}_{n,i}^{AL} \neq 0$ for n large. Therefore, since $\lambda_{n,i}^* = O_p(1)$ and $\lambda_n/\sqrt{n} = o_P(1)$, it turns out that $\lambda_n \lambda_{n,i}^* \left[|\hat{\beta}_{n,i}^{LS} + u_i/\sqrt{n}| - |\hat{\beta}_{n,i}^{LS})| \right] = o_P(1)$. On the the other hand, if $\beta_i = 0$, then $\hat{\beta}_{n,i}^{AL} = 0$ for n large. Furthermore, in this case $\lambda_{n,i}^* = O_p(\sqrt{n})$. Therefore, the limit $R^{AL}(u)$ of $\tilde{R}_n^{AL*}(u)$ is given by

$$
R^{AL}(u) = \begin{cases} -2u_A' W^A + u_A' C^A u_A & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty & \text{otherwise,} \end{cases}
$$

where $W^A \sim N(0, \Omega^A)$, and Ω^A is the sub-matrix of Ω for the nonzero coefficients. Note that the unique minimum of $R^{AL}(u)$ is $((C^A)^{-1} W^A, 0)'$. Therefore, using the results in Geyer (1994) Theorem 3.1 is established. $\square$

Theorem 3.1 shows that the modified pairs bootstrap provides a valid approach for approximating the sampling distribution of T_n^{AL}. Furthermore, by extending the results in Corollary 4.2 and 4.3 in Chatterjee and Lahiri (2011) to our setting, we can show that the modified pairs bootstrap can also be applied for the implementation of testing hypotheses on the regression parameter, and the variance estimation of the adaptive Lasso estimator.

Also in this setting, the implementation of the modified pairs bootstrap differs considerably from the residual bootstrap method proposed in Chatterjee and Lahiri (2011), and the wild bootstrap procedure introduced in Minnier, Tian and Cai (2011). This result shows that when properly defined also the pairs bootstrap may provide a valid approach for estimating the distribution of the adaptive Lasso estimator.

Remark 5. In this remark, we better clarify the failure of the naive pairs bootstrap for the adaptive Lasso estimator, and the validity of the modified pairs bootstrap. First, note that also in this setting $E^*[(y_t^* - x_t'^* \hat{\beta}_n^{AL})x_t^*] \neq 0$. Therefore, to overcome this distortion it is necessary to recenter the bootstrap statistic with respect to the least squares estimator $\hat{\beta}_n^{LS}$. Indeed, the least squares estimator satisfies $E^*[(y_t^* - x_t'^* \hat{\beta}_n^{LS})x_t^*] = 0$, and corresponds to the "true" value for the random bootstrap samples. Finally, because of this recentering it is also necessary to adjust the bootstrap adaptive Lasso estimation criterion using the term $\hat{\beta}_{n,i}^{LS} \cdot \mathbb{I}(\hat{\beta}_{n,i}^{AL} = 0)$. Indeed, with this adjustment the bootstrap adaptive Lasso estimation criterion shrinks the zero coefficients of β to $\hat{\beta}_{n,i}^{LS}$ (the "true" value for the bootstrap samples). This adjustment exactly mimics the (standard) adaptive Lasso estimation criterion that shrinks the zero coefficients of β to 0 (the "true" value for the original sample).

Remark 6. Also in this setting, the accuracy of the adaptive Lasso estimator may heavily depend on the selection of the tuning parameter λ_n. Using the results in Theorem 3.1, we can extend the data-driven method proposed in Remark 4 also to adaptive Lasso estimators. More precisely, for $\lambda_n \in [0, +\infty)$ let $\tilde{\beta}_n^{AL*} = \tilde{\beta}_n^{AL*}(\lambda_n)$ denote the bootstrap modified adaptive Lasso estimator. Using Theorem 3.1, we can estimate the mean squared error $E[\|\hat{\beta}_n^{AL} - \beta\|^2]$ of $\hat{\beta}_n^{AL}$ by

$$\phi(\lambda_n) = E^*[\|\tilde{\beta}_n^{AL}(\lambda_n) - \hat{\beta}_n^{LS}\|^2]. \tag{8}$$

Finally, we select the optimal value $\hat{\lambda}_n$ that minimizes the estimated mean squared error (8). In the next section, we study through Monte Carlo simulations the accuracy of the modified pairs bootstrap combined with this data-driven method.

Remark 7. For the sake of brevity, in this paper we focus only on the Lasso (Section 2) and the adaptive Lasso (Section 3) estimators. However, in principle the modified pairs bootstrap could also be applied to a wider class of penalized regression estimators. More precisely, in order to define the modified pairs bootstrap we need two corrections: (i) the bootstrap statistic has to be recentered with respect to the least squares estimator, and (ii) the bootstrap penalization term has to be properly adjusted.

Remark 8. In our analysis, we consider settings where the regression parameter β is fixed, and the sample size n is large. As pointed out in Chatterjee and Lahiri (2011), models with fixed regression parameter are of interest in many applications, and the estimation of the distribution of Lasso estimators is a very important issue. The results in this paper show that the modified pairs bootstrap overcomes indeed this problem.

As highlighted in Remarks 2 and 5, the modified pairs bootstrap recenters the bootstrap statistic with respect to the least squares estimator, and consequently relies on the consistency of $\hat{\beta}_n^{LS}$. Therefore, using the results in Huber and Ronchetti (2009) and by imposing appropriate assumptions on the linear regression model (1), it may be possible to extend the findings in this paper also to settings where the dimension $p < n$ of the regression parameter β is allowed to depend on the sample size n (see, e.g., Chapter 7 in Huber and Ronchetti (2009) and references therein for the consistency and asymptotic normality of the least squares estimator in linear regression models where p is permitted to grow). On the other hand, the (possible) extension and definition of modified pairs bootstrap procedures to more general high-dimensional settings where the dimension of the regression parameter may be larger than n remains unclear and is currently under investigation by the author.

4 Monte Carlo Simulations

We now illustrate the theoretical findings obtained in the previous sections by simulation. We consider the linear regression model (1) with $p = 10$, $\beta = (2, 2, 1, 1, 0.5, 0.5, 0, 0, 0, 0)$, and $n = 200, 400$. Furthermore, for the covariates and the error terms we consider three different settings: (i) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, 1)$, (ii) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p}\sum_{i=1}^p |x_{t,i}|$ and (iii) $x_{t,i} \sim \chi_1^2$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p}\sum_{i=1}^p x_{t,i}$. As in Chatterjee and Lahiri (2011), we study the finite sample performance of bootstrap confidence sets for the regression parameter β. In the first exercise, we consider the Lasso estimator. For each setting, we generate $N = 3,000$ random samples, and for each sample we construct confidence sets for the parameter β of the form

$$I_{n,\alpha}^{L,K} = \{t \in \mathbb{R}^p : \|t - \hat{\beta}_n^L\| \leq t_n^{L,K}(\alpha)\}, \tag{9}$$

where $K = NB, MB$, $t_n^{L,NB}(\alpha)$ is the α-quantile of the naive pairs bootstrap distribution of $\|\hat{\beta}_n^{L*} - \hat{\beta}_n^L\|$, $t_n^{L,MB}(\alpha)$ is the α-quantile of the modified pairs bootstrap distribution of $\|\tilde{\beta}_n^{L*} - \hat{\beta}_n^{LS}\|$, and $\alpha = 0.9, 0.95, 0.99$ is the nominal coverage probability. We construct the bootstrap distributions based on $B = 299$ bootstrap replications. For the modified pairs bootstrap, we select the tuning parameters $\lambda_n \in [0, 10\sqrt{n}]$ and $a_n \in (0, 0.5)$ according to the data-drive method introduced in Remark 4. Similarly, for the naive pairs bootstrap we select the tuning parameter $\lambda_n \in [0, 10\sqrt{n}]$ by minimizing the estimated mean squared error of $\hat{\beta}_n^L$ based on the naive pairs bootstrap approximation.

In Table 1, we report the empirical coverages. We can observe that the modified pairs bootstrap always provides empirical coverages very close to the nominal coverage probability. For instance, in the top panel of Table 1 for $n = 200$, the empirical coverages for the modified pairs bootstrap are 0.8907, 0.9427, and 0.9817, for $\alpha = 0.9$, $\alpha = 0.95$ and $\alpha = 0.99$, respectively. Furthermore, also in the second panel of Table 1 for $n = 400$, the difference between modified pairs bootstrap empirical coverages and nominal coverage probability is always smaller than 0.01. In contrast, the empirical coverages using the naive pairs bootstrap tend to be dramatically far from the nominal coverage probability. For instance, in the top panel of Table 1 for $n = 400$, the naive pairs bootstrap empirical coverages are 0.0210, 0.0440 and 0.1447, for $\alpha = 0.9$, $\alpha = 0.95$ and $\alpha = 0.99$, respectively. To better evaluate the inaccuracy of the naive pairs bootstrap approximation, note that the Lasso estimation criterion shrinks to 0 both the Lasso estimator $\hat{\beta}_n^L$ and the bootstrap Lasso estimator $\hat{\beta}_n^{L*}$. Therefore, the naive pairs bootstrap statistic $\|\hat{\beta}_n^{L*} - \hat{\beta}_n^L\|$ is not able to reproduce the target variability of the original statistic $\|\hat{\beta}_n^L - \beta\|$. This explain the strong undercoverage of the naive pairs bootstrap confidence sets. On the other hand, the modified pairs bootstrap distribution of $\|\tilde{\beta}_n^{L*} - \hat{\beta}_n^{LS}\|$ provides a valid and accurate approximation of the sampling distribution of $\|\hat{\beta}_n^L - \beta\|$.

In the second exercise, we study the finite sample performance of bootstrap confidence sets for the regression parameter β using the adaptive Lasso estimator. We consider the same three settings analyzed in the previous exercise. For $K = NB, MB$ we construct confidence sets for the parameter β of the form

$$I_{n,\alpha}^{AL,K} = \{t \in \mathbb{R}^p : \|t - \hat{\beta}_n^{AL}\| \leq t_n^{AL,K}(\alpha)\}, \tag{10}$$

where $t_n^{AL,NB}(\alpha)$ is the α-quantile of the naive pairs bootstrap distribution of $\|\hat{\beta}_n^{AL*} - \hat{\beta}_n^{AL}\|$, $t_n^{AL,MB}(\alpha)$ is the α-quantile of the modified pairs bootstrap distribution of $\|\tilde{\beta}_n^{AL*} - \hat{\beta}_n^{LS}\|$, and $\alpha = 0.9, 0.95, 0.99$ is the nominal coverage probability. We construct the bootstrap distributions based on $B = 299$ bootstrap replications. For the modified pairs bootstrap, we select the tuning parameter $\lambda_n \in [0, 2\sqrt{n}]$ according to the data-drive method introduced in Remark 6. Similarly, for the naive pairs bootstrap we select the tuning parameter $\lambda_n \in [0, 2\sqrt{n}]$ by minimizing the estimated mean squared error of $\hat{\beta}_n^{AL}$ based on the naive pairs bootstrap approximation.

In Table 2, we report the empirical coverages. Also in this case, it is interesting to note that the modified pairs bootstrap always provides empirical coverages very close to the nominal coverage probability. For instance, in the bottom panel of Table 2 for $n = 400$, the empirical coverages for the modified pairs bootstrap are 0.8903, 0.9453, and 0.9877 for $\alpha = 0.9$, $\alpha = 0.95$ and $\alpha = 0.99$, respectively. On the other hand, the empirical coverages using the naive pairs bootstrap tend again to be slightly smaller than the nominal coverage probability. For instance, in the second panel of Table 2 for $n = 400$, the naive pairs bootstrap empirical coverages are 0.7887, 0.8737 and 0.9610 for $\alpha = 0.9$, $\alpha = 0.95$ and $\alpha = 0.99$, respectively. In conclusion, also these empirical findings confirm the accuracy of the modified pairs bootstrap.

5 Concluding Remarks

In this paper, we study the validity of the pairs bootstrap for Lasso estimators in linear regression models with random covariates and heteroscedastic error terms. In particular, we provide three main contributions to the existing literature. First, we show that the naive pairs bootstrap may have some issues in approximating the distribution of the Lasso estimator. Second, to overcome this problem we introduce a modified pairs bootstrap procedure that consistently estimates the distribution of the Lasso estimator. Third, we show that the modified pairs bootstrap can also be implemented to consistently estimate the distribution of the adaptive Lasso estimator. In conclusion, these results show that when properly defined the pairs bootstrap may provide a valid approach for estimating the distribution of Lasso estimators.

Besides the consistency of the modified pairs bootstrap, it could be interesting to verify whether this bootstrap approach may also imply some asymptotic refinements. By extending the recent study in Chatterjee and Lahiri (2013) to our setting, we believe that the modified pairs bootstrap (applied to an appropriate asymptotically pivotal statistic) should improve approximations based on the limit distribution. This analysis is left for future research.

References

[1] Audrino, F. and Camponovo, L. (2013). Oracle properties and inference of the adaptive Lasso for time series regression models. Working paper.

[2] Bühlmann, P. and van de Geer, S. (2011). Statistics for high-dimensional data: Methods, theory and applications. Springer Series in Statistics, Springer, Heidelberg.

[3] Chatterjee, A. and Lahiri, S.N. (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.*, **138**, 4497-4509.

[4] Chatterjee, A. and Lahiri, S.N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, **106**, 608-625.

[5] 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. *Annals of Statistics*, **41**, 1232-1259.

[6] Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, **9**, 1218-1228.

[7] Geyer, C. (1994). On the asymptotics of constrained M-estimation. *Annals of Statistics*, **22**, 1993-2010.

[8] Hall, P., Lee, E.R. and Park, B.U. (2009). Bootstrap-based penalty choice for the lasso achieving oracle performance. *Statistica Sinica*, **19**, 449-471.

[9] Huber, P. J. and Ronchetti, E.M. (2009). Robust Statistics. Wiley, New York.

[10] Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, **28**, 1356-1378.

[11] Minnier, J., Tian, L. and Cai, T. (2011). A perturbation method for inference on regularized regression estimates. *Journal of the American Statistical Association*, **106**, 1371-1382.

[12] Tibshirani, R. (1996). Regression analysis and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288.

[13] Wagener, J. and Dette, H. (2012). Bridge estimators and the adaptive Lasso under heteroscedasticity. *Mathematical Methods of Statistics*, **21**, 109-126.

[14] Wainwright, M. J. (2006). Sharp thresholds for high-dimensional and noisy recovery of sparsity. Tech. rep., Dept. of Statistics, UC Berkeley.

[15] Zhao, P. and Yu, B. (2006). On model consistency of lasso. *Journal of Machine Learning Research*, **7**, 2541-2563.

[16] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418-1429.

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.1417	0.8907
	0.95	0.2287	0.9427
	0.99	0.4833	0.9817
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.0210	0.8890
	0.95	0.0440	0.9463
	0.99	0.1447	0.9877

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.1663	0.8710
	0.95	0.2443	0.9360
	0.99	0.4943	0.9880
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.0343	0.8903
	0.95	0.0573	0.9413
	0.99	0.1647	0.9873

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.7940	0.8610
	0.95	0.8960	0.9273
	0.99	0.9777	0.9837
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.7370	0.8710
	0.95	0.8547	0.9363
	0.99	0.9663	0.9893

Table 1: **Lasso Estimator: Bootstrap Empirical Coverages.** We report the naive pairs bootstrap and the modified pairs bootstrap empirical coverages for the linear regression model (1) with $p = 10$, $\beta = (2, 2, 1, 1, 0.5, 0.5, 0, 0, 0, 0)$, and $n = 200, 400$. From the top to the bottom panels, we consider three different settings: (i) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, 1)$, (ii) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p}\sum_{i=1}^{p} |x_{t,i}|$ and (iii) $x_{t,i} \sim \chi_1^2$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p}\sum_{i=1}^{p} x_{t,i}$. The nominal coverage probability is $\alpha = 0.90, 0.95, 0.99$.

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.8360	0.9120
	0.95	0.9187	0.9623
	0.99	0.9723	0.9923
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.8140	0.9110
	0.95	0.8970	0.9580
	0.99	0.9710	0.9917

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.8003	0.8990
	0.95	0.8793	0.9520
	0.99	0.9640	0.9933
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.7887	0.9063
	0.95	0.8737	0.9510
	0.99	0.9610	0.9923

$n = 200$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.8700	0.8807
	0.95	0.9320	0.9473
	0.99	0.9853	0.9927
$n = 400$	α	Naive Pairs Bootstrap	Modified Pairs Bootstrap
	0.90	0.8657	0.8903
	0.95	0.9220	0.9453
	0.99	0.9793	0.9877

Table 2: **Adaptive Lasso Estimator: Bootstrap Empirical Coverages.** We report the naive pairs bootstrap and the modified pairs bootstrap empirical coverages for the linear regression model (1) with $p = 10$, $\beta = (2, 2, 1, 1, 0.5, 0.5, 0, 0, 0, 0)$, and $n = 200, 400$. From the top to the bottom panels, we consider three different settings: (i) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, 1)$, (ii) $x_{t,i} \sim N(0, 1)$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p} \sum_{i=1}^{p} |x_{t,i}|$ and (iii) $x_{t,i} \sim \chi_1^2$ and $\epsilon_t \sim N(0, \sigma_t^2)$, where $\sigma_t = \frac{1}{p} \sum_{i=1}^{p} x_{t,i}$. The nominal coverage probability is $\alpha = 0.90, 0.95, 0.99$.