Statistica Sinica **10**(2000), 497-515

VARIANCE ESTIMATION IN HIGH DIMENSIONAL REGRESSION MODELS

Snigdhansu Chatterjee and Arup Bose

Indian Statistical Institute, Calcutta

Abstract: We treat the problem of variance estimation of the least squares estimate of the parameter in high dimensional linear regression models by using the Uncorrelated Weights Bootstrap (UBS). We find a representation of the UBS dispersion matrix and show that the bootstrap estimator is consistent if $p^2/n \to 0$ where p is the dimension of the parameter and n is the sample size. For fixed dimension we show that the UBS belongs to the R-class as defined in Liu and Singh (1992).

Key words and phrases: Bootstrap, dimension asymptotics, jackknife, many parameter regression, variance estimation.

1. Introduction

In Efron (1979) the *bootstrap* method was introduced to understand the *jackknife* better, and it is a general technique to estimate the distribution of statistical functionals. Broadly, the bootstrap principle is to sample from the data itself with replacement and to compute the statistic for each such resample, and appropriately average over all possible resamples. This can be viewed as attaching a random weight to each data point, computing the statistic for the randomised data and then integrating out the extraneous randomisation. Sampling from the data with replacement is thus attaching $Multinomial(n; 1/n/1/n, \ldots, 1/n)$ weights to the n data points. Other random weights, satisfying certain sets of conditions, can also be used for resampling. Any such resampling technique may be called a *generalised bootstrap*.

The idea of bootstrapping with random weights probably appeared first in Rubin (1981). Bootstrapping with exchangeable weights have been treated in Efron (1982), Lo (1987), Weng (1989), Zheng and Tu (1988), Praestgaard and Wellner (1993). Other generalised bootstrap methods may be found in Boos and Monahan (1986), Lo (1991), Härdle and Marron (1991), Mammen (1993). A review can be found in Barbe and Bartail (1995). In this paper we focus on estimating the mean squared error of the least squares estimator of the regression parameter in high dimensional linear models by using a generalised bootstrap technique.

The data consists of the observations $\{(\mathbf{x}_{i:n}, y_{i:n}), i = 1, \ldots, n\}$, where $\mathbf{x}_{i:n}$'s are $p_n \times 1$ vectors and $y_{i:n}$'s are real valued. We assume the linear regression model given by

$$y_{i:n} = \mathbf{x}_{i:n}^T \beta_n + e_{i:n}, \; i = 1, \ldots, n, \tag{1.1}$$

for some fixed but unknown sequence of constants β_n and some function $e_{i:n}$ which represents the "error". We henceforth write $p, \beta, \mathbf{x}_i, y_i, e_i$ respectively for $p_n, \beta_n, \mathbf{x}_{i:n}, y_{i:n}, e_{i:n}$. Let $\mathbf{X}$ denote the $n \times p$ matrix whose ith row is formed by $\mathbf{x}_i^T$. Let $\mathbf{X}^T$ denote the transpose of $\mathbf{X}$. Also let $\mathbf{Y}$ and $\mathbf{e}$ be the n-dimensional vectors whose ith entries are y_i and e_i respectively. Then the model with n observations may be written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}. \tag{1.2}$$

Let $\mathbf{P_x} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ be the projection matrix on the column space of $\mathbf{X}$. Let $\mathbf{A}_{ij}$ denote the (i, j)th element of the matrix $\mathbf{A}$. Also, for any real symmetric matrix $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$, $\lambda_{\mathrm{amax}}(\mathbf{A})$, $\lambda_i(\mathbf{A})$, respectively, denote the maximum, minimum, maximum in absolute value and ith eigenvalue of $\mathbf{A}$. Throughout the rest of the paper we use the generic c and k to denote constants, without implying they are the same, wherever they appear.

We now state the conditions which we impose on our linear model. The regressors $\mathbf{x}_i$'s may or may not be random. The first condition, on the dimension growth, is

$$p^2/n \to 0 \text{ as } n \to \infty. \tag{1.3}$$

We assume the following conditions if $\mathbf{x}_i$'s are non-random:

$$\sup_{1 \le i \le n} ||\mathbf{x}_i||^2 = O(p), \tag{1.4}$$

$$\lambda_{\min}(n^{-1}\mathbf{X}^T\mathbf{X}) > c > 0. \tag{1.5}$$

The assumptions on the errors are

$$Ee_i^2 = \tau_i^2 < c < \infty, \tag{1.6}$$

$$Ee_ie_j = 0; \quad i, j \text{ different}, \tag{1.7}$$

$$\sup Ee_i^2 e_j e_k = O(n^{-1}), \; i, j, k \text{ different}, \tag{1.8}$$

$$\sup Ee_ie_je_ke_l = O(n^{-2}), \; i, j, k, l \text{ different}, \tag{1.9}$$

$$\sup Ee_i^4 < \infty. \tag{1.10}$$

In case the $\mathbf{x}_i$'s are random, we need probabilistic versions of (1.4) and (1.5). Suppose $\mathcal{A}$ is the set on which $\lambda_{\min}(n^{-1}\mathbf{X}^T\mathbf{X}) > c > 0$. Then the conditions assumed are

$$\sup_{1\leq i\leq n} E||\mathbf{x}_i||^2 = O(p), \tag{1.11}$$

$$P[\mathcal{A}] = 1 - O(p^2n^{-2}). \tag{1.12}$$

When $\mathbf{x}_i$'s are random, the assumptions on the error are

$$e_i \text{ is independent of } \mathbf{x}_i, (\mathbf{x}_j, e_j),\ j \leq i \quad \forall i, \tag{1.13}$$

$$Ee_i = 0, \tag{1.14}$$

$$Ee_i^2 = \tau_i^2 < c < \infty, \tag{1.15}$$

$$\sup Ee_i^4 < \infty. \tag{1.16}$$

A precise definition of our models is now as follows:

Model 1. (Fixed regressors) This is (1.1) with non-random $\mathbf{x}_i$'s satisfying conditions (1.3) - (1.5) and (1.6)-(1.10).

Model 2. (Random regressors) This is (1.1) with random $\mathbf{x}_i$'s satisfying conditions (1.3), (1.11)-(1.12) and (1.13)-(1.16).

It may be mentioned here that our conditions on the errors allow for the e_i's to come from several standard dependent structures, such as the autoregressive or autoregressive conditional heteroscedastic. If the e_i's are mean zero normal random variables, not necessarily independent, then (1.8) and (1.9) follow from (1.7). This follows from

Lemma 1.1. (Wick) *If* (N_1, N_2, N_3, N_4) *is a normal random vector with mean zero then*

$$E(N_1N_2N_3N_4) = E(N_1N_2)E(N_3N_4) + E(N_1N_3)E(N_2N_4) + E(N_1N_4)E(N_2N_3).$$

Bootstrap schemes for linear models have been discussed in Efron (1979), Freedman (1981) and in Bickel and Freedman (1983). In linear models, generalised bootstrap may be performed in essentially two ways: by resampling the residuals after fitting the model, or by resampling the data pairs $\{(\mathbf{x}_i, y_i),\ i = 1,\ldots,n\}$. If multinomial weights are used, then these resampling schemes are usually known as the residual bootstrap and the paired bootstrap respectively. Hinkley (1977), Wu (1986), Shao and Wu (1987) have studied consistency of different bootstrap and jackknife schemes in heteroskedastic linear models. The bootstrap in regression models with many parameters has been considered by Bickel and Freedman (1983) and Mammen (1993), who respectively showed the consistency of the residual bootstrap and the wild bootstrap (which is a generalisation of the residual bootstrap) and paired bootstrap for the distribution of least squares estimate of the regression parameters. A generalised bootstrap which uses the pairs for resampling does not seem to have received enough attention. We focus on such resampling schemes here for estimating the mean squared error of the least squares estimator. The resampling scheme is carried out by weighting each data point $(y_i, \mathbf{x}_i)$ with the random weight $\sqrt{w_i}$, then computing the statistic of interest and taking expectation of the random weight vector. In particular, this generalised bootstrap includes the paired bootstrap and all the delete-d jackknives. We call our scheme the *uncorrelated bootstrap* (UBS) and the precise conditions on the weights are given in the next section.

Our generalised bootstrap can be looked upon as a weighted least squares analysis with random weights, and then simple averaging over such weighted least squares estimates. With easily available weighted least squares routines this can lead to significant ease in implementing resampling in linear models. Even though we discuss estimating the variance of the ordinary least squares estimator here, our technique potentially extends to the corresponding problem for the weighted least squares.

Note that (1.5) or (1.12) ensures that the inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ exists. Thus we want to estimate $\mathrm{V_n} = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T$. Under Model 1, this is $\mathrm{V_n} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$, where $\mathbf{T}$ is a diagonal matrix with ith diagonal element τ_i^2. Notice that this is a $p \times p$ matrix, and for us $p \to \infty$. Hence we try to estimate $\xi^T \mathrm{V_n} \xi$ for any $\xi \in \mathbb{R}^p$ with $||\xi|| = 1$. Our main result is a representation for $\xi^T \mathrm{V_{UBS}} \xi$, where $\mathrm{V_{UBS}}$ is an appropriate bootstrap estimate of $\mathrm{V_n}$. This also gives an element wise bound of $\mathrm{V_n}$ and results for estimating the variance of linear combinations of the elements of β.

For linear models with nonrandom design and fixed p, Liu and Singh (1992) studied bootstrap and jackknife schemes. They showed that for estimating the variance of the least squares estimate of β, some resampling schemes such as the paired bootstrap and wild bootstrap produce consistent results under heteroskedasticity, while some others such as the usual residual bootstrap do not yield consistent estimates under heteroskedasticity but are more efficient under homoskedasticity. These resampling techniques are thus either robust or efficient: they belong to the R-class or the E-class. Our results show that for fixed p the UBS we study belongs to the R-class. Note that two special cases of UBS, the paired bootstrap and the delete-1 jackknife were already known to belong to the R-class.

The random regressor model has been considered in Mammen (1993) in the context of the paired bootstrap and the wild bootstrap. The model there was based on observing *i.i.d.* variables $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, n\}$, where $\mathbf{x}_i$'s were $p \times 1$

vectors with the assumed model relation being $y_i = \mathbf{x}_i^T\beta + e_i$ for some constant but unknown β. The dimension p was allowed to vary with n. Note that this implies that $\{(\mathbf{x}_i, e_i),\ i = 1, \ldots, n\}$ are also *i.i.d.* and this, in turn, implies (1.13)-(1.15). Chatterjee and Bose (1998) considered the problem of estimating the distribution of the least squares estimate of a linear regression parameter, using the UBS resampling scheme, in the same set-up as that of Mammen (1993). It was shown that the distribution of any contrast of the least squares estimator and its UBS bootstrap equivalent tend to the same normal limit, thus establishing consistency of the UBS technique for estimating the distribution function. Since we are estimating the variance of the least squares estimator as opposed to the distribution function, as in Mammen (1993), the fourth moment condition (1.16) is imposed.

Our random regressor model is more general than Mammen (1993) in the sense that the assumption of *i.i.d.* nature of the data is modified to (1.13). However, in place of assuming $p^{1+\delta}/n \to 0$ for $\delta \geq 1/3$ as in Mammen (1993), or $\delta > 0$ as in Chatterjee and Bose (1998), we need $p^2/n \to 0$. This is explained by the fact that in general the bias in the least squares estimate is of the order $p/n^{1/2}$, and so the mean squared error that we are estimating requires (1.3).

We now check that with regressors taken to be independently and identically distributed, as assumed in Mammen (1993), our model conditions (1.11)-(1.12) hold. Suppose the regressors $\mathbf{x}_{i:n}$ are *i.i.d.*, with $\sup_n \sup_{||d||=1} E|d^T\mathbf{x}_{i:n}|^4 < \infty$ as in condition 2.2 of Theorem 1 of Mammen (1993). The condition (1.11) follows directly from Lemma 0 of Mammen (1993). Assume without loss that $E\mathbf{x}_i\mathbf{x}_i^T = \mathbf{I}$, and letting $\mathbf{A} = n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T - \mathbf{I}$, we have $n^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{I} + \mathbf{A}$, so that

$$\lambda_{\min}(n^{-1}\mathbf{X}^T\mathbf{X}) = 1 + \lambda_{\min}(\mathbf{A}) \geq 1 - \lambda_{\text{amax}}(\mathbf{A})$$

and hence

$$\begin{aligned} P[|\lambda_{\min}(n^{-1}\mathbf{X}^T\mathbf{X})| \leq 1/2] &\leq P[|1 - \lambda_{\text{amax}}(\mathbf{A})| \leq 1/2] \\ &\leq P[|\lambda_{\text{amax}}(\mathbf{A})| \geq 1/2] \\ &\leq 4E(\lambda_{\text{amax}}(\mathbf{A}))^2 = O(p^{3/2}n^{-1}), \end{aligned}$$

with the last relation following from Lemma 1 of Mammen (1993). This verifies (1.12).

For some resampling schemes that we consider, condition (1.5) is not sufficient. For example, a paired bootstrap sample can include only the data points indexed by $\mathcal{I}_m = \{i_1, \ldots, i_m\}$. If $m < p$ there is no possibility that the design matrix is of full column rank. However, this case has exponentially small probability as we show later. Even if $m \geq p$, we still need nonsingularity of $\mathbf{X}^{T*}\mathbf{X}^*$. So an equivalent of (1.5) is needed for submatrices of the design matrix.

We now state here a more general condition. Suppose m_0 is a specified integer in the range $[n/3]$ to n. For any integer m in $\{m_0, \ldots, n\}$ consider the subset $\mathcal{I}_m = \{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$. Let $\mathbf{X}^*$ be the $m \times p$ matrix whose jth row is $\mathbf{x}_{i_j}^T$. Then the general condition is

$$m^{-1}\mathbf{X}^{T*}\mathbf{X}^* > k_1\mathbf{I}, \quad k_1 > 0 \tag{1.17}$$

for every such choice of subset $\mathcal{I}_m$ of size m from $\{1, \ldots, n\}$ and for every m in $[m_0, n]$. (For two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ we write $\mathbf{A}_1 > \mathbf{A}_2$ if and only if $\mathbf{A}_1 - \mathbf{A}_2$ is positive definite.) Note that this condition depend on m_0; the higher its value, the weaker is the condition. For $m_0 = n$ condition (1.17) is same as (1.5). For the model with random regressors, the corresponding condition is that the set $\mathcal{A}$, on which

$$m^{-1}\mathbf{X}^{T*}\mathbf{X}^* > k_1\mathbf{I}, \text{ for every } \mathcal{I}_m \text{ and every } m \in [m_0, n], \tag{1.18}$$

has a probability $1 - O(p^2/n)$. There are UBS resampling schemes which require only $m_0 = n$. The delete-d jackknife requires $m_0 = n - d$. However, the paired bootstrap requires as low an m_0 as possible. The more stringent assumption (1.17) or (1.18) is required to make resampling schemes like the paired bootstrap and the different jackknives feasible.

2. The Resampling Scheme

Let $\{w_{i:n}; 1 \leq i \leq n, n \geq 1\}$ be a triangular array of non-negative random variables to be used as weights. *We drop the suffix n from the notation of the weights.* The resampling scheme is carried out by weighting each data point $(y_i, \mathbf{x}_i)$ with the random weight $\sqrt{w_i}$, then computing the statistic of interest and taking expectation of the random weight vector. The above set up can be taken as a direct generalization of the *paired bootstrap*, where the $\{w_i; 1 \leq i \leq n\}$ are given by a random sample from $Multinomial(n; 1/n, \ldots, 1/n)$. The different delete-$d$ jackknives can also be viewed as special cases of this resampling technique, see Chatterjee (1998) for details.

The weights $w_1, \ldots, w_n$ used for resampling satisfy certain restrictions on the first few moments which we now state. Let $V(w_i) = \sigma_n^2$ and assume that the quantities

$$E\Big(\frac{w_a - 1}{\sigma_n}\Big)^i\Big(\frac{w_b - 1}{\sigma_n}\Big)^j\Big(\frac{w_c - 1}{\sigma_n}\Big)^k \ldots$$

are functions of the powers $i, j, k \ldots$ only, and not of the indices $a, b, c \ldots$. Thus we can write

$$c_{ijk\ldots} = E\Big(\frac{w_a - 1}{\sigma_n}\Big)^i\Big(\frac{w_b - 1}{\sigma_n}\Big)^j\Big(\frac{w_c - 1}{\sigma_n}\Big)^k \ldots$$

Note that if the weights are assumed to be exchangeable, then the above condition follows. But exchangeability of weights is not a necessary condition.

Let $\mathcal{W}$ be the set on which at least m_0 of the weights are greater than some fixed constant $k_2 > 0$. The value of m_0 is the same as that of assumptions (1.17) or (1.18). The weights are assumed to satisfy certain conditions.

$$E(w_i) = 1, \tag{2.1}$$

$$\sigma_n^2 \to k > 0, \tag{2.2}$$

$$\mathrm{P_B}[Kn \geq \sum_{i=1}^{n} w_i \geq kn,\ K > k > 0] = 1, \tag{2.3}$$

$$\mathrm{P_B}[\mathcal{W}] = 1 - O(p^2 n^{-1}), \tag{2.4}$$

$$c_{11} = O(n^{-1}), \tag{2.5}$$

$$c_{i_1 \ldots i_k} = O(n^{-k+1})\ \forall\ i_1, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 3, \tag{2.6}$$

$$c_{i_1 \ldots i_k} = O(\min(n^{-k+2}, 1))\ \forall\ i_1, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 4. \tag{2.7}$$

We define the bootstrap estimate of β to be

$$\hat{\beta}_{\mathrm{B}} = \begin{cases} (\mathbf{X}^T \mathbf{W}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} & \text{on the set } \mathcal{W} \cap \mathcal{A}, \\ \hat{\beta} & \text{otherwise,} \end{cases}$$

and the bootstrap variance estimate to be

$$\mathrm{V_{UBS}} = \sigma_n^{-2} \mathrm{E_B} (\hat{\beta}_{\mathrm{B}} - \hat{\beta})(\hat{\beta}_{\mathrm{B}} - \hat{\beta})^T.$$

For Model 1, take the set $\mathcal{A}$ to be the entire sample space, so that the definition of $\hat{\beta}_{\mathrm{B}}$ does not depend on the model used.

If y_i's are *i.i.d.*, the UBS variance estimate coincides with the generalised bootstrap variance estimate used for estimating the variance of the sample mean. See Barbe and Bertail (1995) for use of this statistic for other statistical functionals.

The notable feature of this resampling scheme is that the weights $\{w_{i:n}\}$ are asymptotically uncorrelated. We call this the Uncorrelated Weights Bootstrap (hereafter UBS). We denote the variance estimate in UBS by $\mathrm{V_{UBS}}$. A slight variant of the above conditions on weights can be effected by dropping (2.2) and letting the common variance go to zero, so that the weights are asymptotically degenerate. With that condition, the conditions on various mixed moments can be slightly relaxed. For details about this variation refer to Chatterjee (1998).

One significant condition on the weights is (2.4), which we now discuss. This condition is related to (1.17) and (1.18). If instead, we restrict the condition to require that all the weights are bounded away from zero, then we can take $m_0 = n$ in (1.17). For example, the weights can be taken to be *i.i.d.* unit mean random variables that are supported on some interval (k, K), with $0 < k < K < \infty$. For the delete-d jackknives, we need (1.17) with $m_0 = n - d$ only. For the paired bootstrap, that is for a random sample from $Multinomial(n; 1/n, \ldots, 1/n)$, Proposition 3.1 of the next section shows that (2.4) is satisfied for $m_0 = [n/3]$, the greatest integer less than or equal to $n/3$. Note that the condition on $\mathcal{W}$ is in terms of "at least m_0" weights being positive, so an upper bound on m_0 is really meaningful. There is a duality between model conditions and conditions on the resampling weights. We may relax certain model conditions by making the conditions on resampling weights more stringent. For example, for resampling with independent weights, we can ignore the model conditions (1.8) and (1.9).

3. Main Results

Let

$$(\mathbf{T}_n)_{ij} = \begin{cases} \mathbf{e}_i^2, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Theorem 3.1. *Assume the conditions of Model* 1. *Also assume* (1.17) *for some* m_0. *Then for any* $\xi \in \mathbb{R}^p$ *with* $||\xi|| = 1$,

$$n^{3/2}p^{-1}\xi^T(\mathrm{V_{UBS}} - \mathrm{V_n})\xi = n^{3/2}p^{-1}\xi^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\mathbf{T}_n - \mathbf{T}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\xi + O_P(pn^{-1/2}). \tag{3.1}$$

In particular, UBS is a consistent resampling technique for the model. The distributional asymptotics for the variance estimator can essentially be developed from here, by noting that the leading term in the variance representation is a linear combination of e_i^2's. Note that the leading term in (3.1) is bounded in probability. Since this result does not depend on the particular choice of weights, they can be chosen according to convenience. This exact rate is not obtained under Model 2, but otherwise a very similar representation theorem holds.

Theorem 3.2. *Assume the conditions of Model* 2. *Also assume* (1.18) *for some* m_0. *Then for any* $\xi \in \mathbb{R}^p$ *with* $||\xi|| = 1$,

$$n^{3/2}p^{-1}\xi^T(\mathrm{V_{UBS}} - \mathrm{V_n})\xi = n^{3/2}p^{-1}\xi^T[I_{\mathcal{A}}(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T - \mathrm{V_n}]\xi + O_P(pn^{-1/2})). \tag{3.2}$$

Notice that typically $(\hat{\beta} - \beta) = O_P((n/p)^{-1/2})$. The leading term in (3.2) is of the form $n^{-1/2}\sum_{i=1}^n (Z_i - EZ_i)$ for some standardized random variables Z_i (see (5.7) in the section on proofs for details on the random variables Z_i) plus some remaining terms which are $O_P(pn^{-1/2})$. Suppose we assume that

$E[\xi^T(\hat{\beta}-\beta)]^4 < \infty$. This simultaneously ensures that the indicator on the set $\mathcal{A}$ can be ignored, and that the variance of $(n/p)^{-1/2}(\hat{\beta}-\beta)$, that is $np^{-1}\xi^T \mathrm{V_n}\xi$, is consistently estimated by $np^{-1}\xi^T \mathrm{V_{UBS}}\xi$ in the sense that $np^{-1}\xi^T(\mathrm{V_{UBS}}-\mathrm{V_n})\xi \to 0$ in probability.

Chatterjee (1998) has shown that the conditions on the weights are satisfied by the delete-d jackknives. If $\{w_1,\ldots,w_n\}$ is taken to be a random sample from $Multinomial(n;1/n,\ldots,1/n)$, we get the paired bootstrap. The moment conditions on these weights can be easily verified by direct calculation. Through the following proposition we show that condition (2.4) is also satisfied with $m_0 = n/3$. One important advantage of using a generalised bootstrap scheme in place of paired bootstrap or the jackknives is that calculations may be simpler and faster with undiminished accuracy.

Proposition 3.1. *Suppose* $(X_1,\ldots,X_n)$ *is* $Multinomial(n;1/n,\ldots,1/n)$*. If* $\{m_n\}$ *is such that* $m_n/n < 1/3$*, then the probability that at least* m_n *of the* X*'s are positive is greater than* $1-e^{-\alpha n}$*, for some constant* $\alpha > 0$*.*

4. Some Simulation Results

To gauge how the different UBS schemes perform, especially when we have random regressors and/or dependent errors, we carried out a small simulation experiment. We chose the following five UBS schemes:

(a) the $Multinomial(n;1/n,\ldots,1/n)$ bootstrap(MB);
(b) the $Dirichlet(n;1/n,\ldots,1/n)$ bootstrap(DB);
(c) the $Uniform(1/2,3/2)$ bootstrap(UB);
(d) the $Beta(2,7)$ bootstrap(BB1);
(e) the $Beta(7,2)$ bootstrap(BB2).

Note that the first choice corresponds to the case of the paired bootstrap and the second one corresponds to the case of the Bayesian bootstrap with a Dirichlet process prior. For the last three choices, a sample of size n is generated from the given distribution and used for resampling. The last two schemes were considered to see how asymmetry of the generating distribution affects the performance of the resampling method. Computationally, the third choice is the easiest and fastest.

We considered three models with $p = 1$. The first model is the simple one of estimating the variance of the sample mean. This is intended to serve as a benchmark. The second model is the autoregressive (AR) model of order one with IID innovations. This is a well-known model in time series and the results obtained for this model are an indication of what is to be expected in similar models such as the AR of order $p > 1$. The third model is also the autoregressive model but here the errors, instead of being IID are assumed to have an $ARCH$ (autoregressive conditional heteroscedastic) structure. The $ARCH$

model is widely used in econometrics to model financial time series data, see Bera and Higgins (1993) for a review.

We now give a precise description of the three models.

Experiment 1. $\mathbf{X}_t = \beta + \epsilon_t$. In this case $\hat{\beta} = n^{-1}\sum_{t=1}^{n}\mathbf{X}_t/n$ and $\mathrm{V_n} = n^{-1}$. For simulations we fix $\beta = 7.0$ and take the errors ϵ_t to be an *i.i.d.* sequence from $Normal(0,1)$. This is the simplest example when Model 1 conditions are satisfied.

Experiment 2. $\mathbf{X}_t = \beta\mathbf{X}_{t-1} + \epsilon_t$. and $\mathbf{X}_0 = 0$. Here $\hat{\beta} = (\sum_{i=1}^{n}\mathbf{X}_{i-1}^2)^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_{i-1}$ and it is known that if $|\beta| < 1$ then $n^{1/2}(\hat{\beta} - \beta) \to N(0, 1-\beta^2)$. For simulation, the innovations ϵ_t are taken to be *i.i.d.* $Normal(0,1)$. The process has an explosive behaviour if $|\beta| > 1$, in which case normal limits are not obtained for the least squares estimate. With this in mind, we choose two different values of β as follows.

Experiment 2(a). We take $\beta = -0.5$. Note that this is well within the safe zone.

Experiment 2(b). We take $\beta = 0.9$. This value is close to the boundary and here the usual bootstrap is not expected to perform well.

Experiment 3. $\mathbf{X}_t = \beta\mathbf{X}_{t-1} + \epsilon_t$. where $\mathbf{X}_0 = 0$ and ϵ_t is an $ARCH$ process satisfying $\epsilon_1 \sim N(0,\gamma^2)$ and $[\epsilon_t|\epsilon_s, s \le t] \sim N(0,\gamma_t^2)$ where γ_t is in general a polynomial in $\epsilon_s, s \le t$. We choose $\gamma_t^2 = \alpha + \beta\epsilon_{t-1}^2$. In this case, the (conditional) least squares estimate of β is given by $\hat{\beta} = (\sum_{i=1}^{n}\mathbf{X}_{i-1}^2)^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_{i-1}$. If $\gamma^2 = \alpha/(1-\beta)$, then it is known that $n^{1/2}(\hat{\beta} - \beta) \to N(0, \gamma^2(1-\beta^2))$. For our simulation we take $\alpha = 0.5$, $\beta = 0.4$ and $\gamma^2 = 5/6$. As earlier, we work with two different values of β. In **Experiment 3(a)** we take $\beta = -0.5$ and in **Experiment 3(b)** we take $\beta = 0.9$.

In each case we fix n, the size of the data, then randomly generate a data set satisfying the assumed conditions and use resampling on it. In all three experiments we first compute the least squares estimate of β, and then resample for the variance of this estimator. Since in all experiments we consider, the least squares estimator has a limiting normal distribution, the different bootstrap variance estimates are compared with the appropriate asymptotic variance.

Results are presented in Tables 1, 2, 3, 4 and 5 respectively. We use the notation *av* for the asymptotic variance. The resample size was 5000 for Experiment 1 and 10000 for Experiment 2(a), 2(b), 3(a), 3(b). As the table entries show, UBS using $Beta(2,7)$ and $Beta(7,2)$ weights lead to slight underestimation and overestimation of the variance, possibly due to the difference in higher order terms. The results from the other resampling schemes are almost identical. UBS with *i.i.d.* uniform weights is recommended.

The starred figures are scaled up 1000 times.

Table 1. Experiment 1.

n	$\hat{\beta}$	$av.*$	MB		DB		UB		BB1		BB2	
			β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$
30	6.88	61.30	6.88	61.92	7.01	80.79	6.88	61.55	6.88	51.99	6.88	82.23
40	6.94	17.99	6.94	17.90	7.00	20.13	6.94	18.14	6.94	15.51	6.95	23.27
100	7.01	10.22	7.01	10.24	6.99	10.93	7.01	10.46	7.01	8.99	7.01	13.21
200	6.99	4.95	6.99	4.90	6.99	5.08	6.99	4.90	6.99	4.30	6.99	6.49
600	6.99	1.66	6.99	1.64	6.99	1.64	6.99	1.67	6.99	1.43	6.99	2.14
1000	6.95	1.10	6.95	1.07	6.95	1.10	6.95	1.11	6.95	0.96	6.95	1.42
2000	7.01	0.47	7.01	0.47	7.01	0.47	7.01	0.48	7.01	0.40	7.01	0.61

Table 2. Experiment 2(a).

n	$\hat{\beta}$	$av.*$	MB		DB		UB		BB1		BB2	
			β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$
30	-0.60	21.40	-0.61	16.44	-0.61	16.19	-0.60	15.32	-0.60	13.83	-0.60	20.32
50	-0.59	13.03	-0.59	12.76	-0.59	11.90	-0.59	12.05	-0.59	10.47	-0.59	15.18
100	-0.52	7.24	-0.53	6.31	-0.52	6.08	-0.52	6.26	-0.53	5.25	-0.53	7.96
200	-0.45	3.99	-0.45	3.71	-0.45	3.54	-0.45	3.70	-0.45	3.16	-0.45	4.76
500	-0.54	1.41	-0.54	1.21	-0.54	1.21	-0.54	1.21	-0.54	1.08	-0.54	1.62
1000	-0.51	0.73	-0.51	0.70	-0.51	0.69	-0.51	0.72	-0.51	0.62	-0.51	0.91
2000	-.50	0.38	-0.50	0.36	-0.50	0.36	-0.50	0.36	-0.50	0.31	-0.50	0.48

Table 3. Experiment 2(b).

n	$\hat{\beta}$	$av.*$	MB		DB		UB		BB1		BB2	
			β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$
30	0.96	2.85	0.93	2.76	0.93	2.66	0.96	3.92	0.96	3.71	0.96	5.50
50	0.92	3.15	0.91	2.66	0.91	2.51	0.92	3.22	0.92	2.79	0.92	4.26
100	0.87	2.49	0.87	2.70	0.87	2.62	0.87	2.81	0.88	2.49	0.88	3.96
200	0.81	1.68	0.81	1.94	0.81	1.87	0.81	1.90	0.81	1.64	0.81	2.51
500	0.85	0.56	0.85	0.58	0.85	0.57	0.85	0.58	0.85	0.50	0.85	0.76
1000	0.91	0.17	0.91	0.17	0.91	0.16	0.91	0.17	0.91	0.15	0.91	0.21
2000	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.09	0.91	0.08	0.91	0.11

Table 4. Experiment 3(a).

n	$\hat{\beta}$	$av.*$	MB		DB		UB		BB1		BB2	
			β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$	β_{B}	$\mathrm{V}^*_{\mathrm{B}}$
30	-0.57	22.39	-0.59	17.12	-0.58	15.52	-0.57	16.65	-0.57	14.19	-0.57	21.78
50	-0.68	10.84	-0.65	36.96	-0.66	27.97	-0.68	47.62	-0.68	39.14	-0.68	61.14
100	-0.68	5.43	-0.67	7.43	-0.67	7.03	-0.68	7.02	-0.68	6.02	-0.68	9.45
200	-0.49	3.80	-0.49	5.43	-0.49	5.25	-0.49	5.37	-0.49	4.81	-0.49	7.14
500	-0.54	1.41	-0.54	1.91	-0.54	1.88	-0.54	1.92	-0.54	1.64	-0.54	2.44
1000	-0.55	0.70	-0.55	2.02	-0.55	1.95	-0.55	2.03	-0.55	1.77	-0.55	2.70
2000	-0.50	0.37	-0.50	0.83	-0.50	0.83	-0.50	0.84	-0.50	0.72	-0.50	1.05

Table 5. Experiment 3(b).

n	$\hat{\beta}$	$av.*$	MB		DB		UB		BB1		BB2	
			$\beta_{\rm B}$	$\rm V_B^*$	$\beta_{\rm B}$	$\rm V_B^*$	$\beta_{\rm B}$	$\rm V_B^*$	$\beta_{\rm B}$	$\rm V_B^*$	$\beta_{\rm B}$	$\rm V_B^*$
30	0.93	4.42	0.91	3.42	0.91	3.35	0.93	5.13	0.93	4.51	0.93	6.76
50	0.94	2.31	0.92	2.84	0.92	2.65	0.94	4.14	0.94	3.54	0.94	5.44
100	0.95	1.05	0.94	1.28	0.94	1.25	0.95	1.68	0.95	1.50	0.95	2.19
200	0.92	0.76	0.92	0.86	0.92	0.84	0.92	0.88	0.92	0.76	0.92	1.15
500	0.91	0.33	0.91	0.38	0.91	0.38	0.91	0.39	0.91	0.35	0.91	0.51
1000	0.90	0.18	0.90	0.24	0.90	0.23	0.90	0.24	0.90	0.21	0.90	0.31
2000	0.88	0.11	0.88	0.16	0.88	0.16	0.88	0.17	0.88	0.14	0.88	0.21

5. Proofs

There are two important conclusions to be derived from the above model conditions. For Model 1 they are

$$\max_i (\mathbf{P}_{\mathbf{x}})_{ii} = O(p/n), \tag{5.1}$$

$$||\sum_{i=1}^{n} \mathbf{x}_i e_i|| = O_P(p^{1/2} n^{1/2}). \tag{5.2}$$

These are quickly verified as follows:

$$\begin{aligned}\max_i (\mathbf{P}_{\mathbf{x}})_{ii} = \max_i \mathbf{x}_i^T (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i &\le n^{-1} \max_i ||\mathbf{x}_i||^2 [\lambda_{\min}(n^{-1}\mathbf{X}^T\mathbf{X})]^{-1} \\ &\le cn^{-1} \max_i ||\mathbf{x}_i||^2 = O(p/n),\end{aligned}$$

and

$$E||\sum_{i=1}^{n} \mathbf{x}_i e_i||^2 = E\sum_{i=1}^{n} e_i^2 \mathbf{x}_i^T \mathbf{x}_i = \sum_{i=1}^{n} \tau_i^2 ||\mathbf{x}_i||^2 \le c \sum_{i=1}^{n} ||\mathbf{x}_i||^2 = O(pn).$$

For Model 2 the equivalent conclusions are

$$\max_i (\mathbf{P}_{\mathbf{x}})_{ii} I_{\mathcal{A}} = O_P(p/n), \tag{5.3}$$

$$||\sum_{i=1}^{n} \mathbf{x}_i e_i|| = O_P(p^{1/2} n^{1/2}). \tag{5.4}$$

These are also easily verified by using similar arguments as earlier.

Note that $\mathbf{X}$ is an $n \times p$ matrix of rank p. Let $\mathbf{X} = \mathbf{PDQ}^T$ be the singular value decomposition of $\mathbf{X}$. That is, $\mathbf{D}$ is a $p \times p$ diagonal matrix with positive diagonal elements, $\mathbf{P}$ is an $n \times p$ matrix such that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ and $\mathbf{Q}$ is a $p \times p$ orthogonal matrix. The spectral representation of $\mathbf{X}^T\mathbf{X}$ is given by $\mathbf{X}^T\mathbf{X} = \mathbf{QD}^2\mathbf{Q}^T$. Let $\Lambda = n^{-1}\mathbf{D}^2$. Note that the minimum eigenvalue of the diagonal matrix Λ is bounded away from zero, so that the maximum eigenvalue of Λ^{-1} is bounded above. The notation $\Lambda^{1/2}$ is sometimes used for $n^{-1/2}\mathbf{D}$. Note that the diagonal entries of $\mathbf{D}$ are all positive.

Let $\mathbf{W}_D$ be the $n \times n$ diagonal matrix with ith diagonal element w_i. Let us also use the notation $W_i = (w_i - 1)/\sigma_n$, and $\mathbf{W}_{(n\times n)} = Diag(w_1, \ldots, w_n)$.

Since our two theorems state almost the identical result under two different models, we use an approach that proves both results simultaneously. Note that under Model 1 the set $\mathcal{A}$ is the entire sample space.

We first prove that with a high probability a condition like (1.5) is also true for the bootstrap design matrix.

Lemma 5.1. *Assume Model* 1. *Also assume condition* (1.17) *for an appropriate choice of $m_0 > p$. Then under the set $\mathcal{W} \cap \mathcal{A}$, all eigenvalues of the matrix $n^{-1}\mathbf{X}^T\mathbf{W}_D\mathbf{X}$ are greater than $k > 0$, where k is a constant.*

Proof of Lemma 5.1. Let $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$ be the eigenvalues of $\mathbf{X}^T\mathbf{W}_D\mathbf{X}$. We have to show a positive lower bound for λ_1/n under $\mathcal{W} \cap \mathcal{A}$. Note that we get λ_1 by minimizing $\xi^T\mathbf{X}^T\mathbf{W}_D\mathbf{X}\xi/\xi^T\xi$ with respect to all vectors $\xi \in \mathbb{R}^p$. Now note that

$$\begin{aligned}\min_{\xi\in\mathbb{R}^p} \frac{\xi^T\mathbf{X}^T\mathbf{W}_D\mathbf{X}\xi}{\xi^T\xi} &= \min_{\xi\in\mathbb{R}^p}\min_{\eta=\mathbf{D}\xi} \frac{\eta^T\mathbf{P}^T\mathbf{W}_D\mathbf{P}\eta}{\eta^T\eta}\frac{\xi^T\mathbf{D}^2\xi}{\xi^T\xi} \\ &\ge \min_{\xi\in\mathbb{R}^p}\min_{\eta\in\mathbb{R}^p} \frac{\eta^T\mathbf{P}^T\mathbf{W}_D\mathbf{P}\eta}{\eta^T\eta}\frac{\xi^T\mathbf{D}^2\xi}{\xi^T\xi}.\end{aligned}$$

By condition (1.5), on the set $\mathcal{A}$ we have n^{-1} times the second term bounded below by a positive constant. Thus in order to complete the proof, we need to show that $\eta^T\mathbf{P}^T\mathbf{W}_D\mathbf{P}\eta/\eta^T\eta$ has a positive lower bound, as η varies. First observe that the model condition (1.17) says that on $\mathcal{A}$, for all $m >$ an appropriate m_0 if $\mathcal{I}_m = \{i_1, \ldots, i_m\}$ is a subset of $\{1, \ldots, n\}$, and if $\mathbf{X}^*$ is the $m \times p$ matrix whose jth row is $\mathbf{x}_{i_j}^T$, then $m^{-1}\mathbf{X}^{T*}\mathbf{X}^* > k_1\mathbf{I}$. If $\mathbf{P}^*$ is the submatrix of $\mathbf{P}$ corresponding to $\mathbf{X}^*$, this implies $\mathbf{P}^{*T}\mathbf{P}^* = \sum_{\{i\in\mathcal{I}_m\}} \mathbf{p}_i\mathbf{p}_i^T > k_1\frac{m}{n}\mathbf{I} \ge \frac{k_1}{3}\mathbf{I}$.

Suppose $\mathcal{S}$ is the set of indices of weights that are greater than some fixed k_2. Under $\mathcal{W}$, $\mathcal{S}$ has m elements, where $m \ge m_0$. Thus the same set of indices $\mathcal{S}$ is also an $\mathcal{I}_m$ for which (1.17) holds. Therefore under the set $\mathcal{W} \cap \mathcal{A}$,

$$\mathbf{P}^T\mathbf{W}_D\mathbf{P} = \sum_{i=1}^n w_i\mathbf{p}_i\mathbf{p}_i^T > k_2 \sum_{\{w_i\in\mathcal{S}\}} \mathbf{p}_i\mathbf{p}_i^T > \frac{k_2k_1}{3}\mathbf{I}.$$

This completes the proof.

Let $\mathbf{U}_B = \mathbf{P}^T\mathbf{W}_D\mathbf{P}$. Then Lemma 5.1 allows us to conclude that

$$\lambda_{\max}(\mathbf{U}_B^{-1})I_{\mathcal{W}\cap\mathcal{A}} < k < \infty. \tag{5.5}$$

Using (5.5) we have a more precise statement about the nature of the matrix $\mathbf{U}_B^{-1}$.

Lemma 5.2. *On the set* $\mathcal{W} \cap \mathcal{A}$

$$\mathbf{U}_B^{-1} = \mathbf{I} + \sigma_n \mathbf{R}_W, \tag{5.6}$$

where $\lambda_{\text{amax}}(\mathrm{E_B}\mathbf{R}_W^2 I_{\mathcal{W}\cap\mathcal{A}}) = O(p^2 n^{-1})$.

Proof of Lemma 5.2. Since $\mathbf{U}_B = \mathbf{P}^T\mathbf{W}_D\mathbf{P}$, we have $\mathbf{U}_B = \mathbf{I} + \sigma_n \mathbf{P}^T\mathbf{W}\mathbf{P} = \mathbf{I} + \sigma_n \mathbf{R}_B$, say. Then it is easily seen that $tr(\mathbf{R}_B^2) = tr(\mathbf{W}\mathbf{P}_\mathbf{x}\mathbf{W}\mathbf{P}_\mathbf{x})$. ¿From (5.1) it now follows that $tr\mathrm{E_B}(\mathbf{R}_B^2) = O(p^2/n)$, thus eventually all the eigenvalues of $\mathrm{E_B}\mathbf{R}_B$ are less than 1, and also $\lambda_{\max}(\mathrm{E_B}\mathbf{R}_B^2) = O(p^2n^{-1})$. This means that by taking an inverse we can write $\mathbf{U}_B^{-1} = \mathbf{I} + \sigma_n \mathbf{R}_B \sum_{k\geq 0} \sigma_n^k \mathbf{R}_B^k = \mathbf{I} + \sigma_n \mathbf{R}_W$, say, then $\lambda_{\max}(\mathrm{E_B}\mathbf{R}_W^2 I_{\mathcal{W}\cap\mathcal{A}}) = O(p^2n^{-1})$.

Proof of Theorem 3.1. Note that

$$\begin{aligned}
&\hat\beta_{\mathrm{B}} - \hat\beta \\
&= [(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}_D\mathbf{e} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]I_{\mathcal{W}\cap\mathcal{A}}\mathbf{e} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{W}_D - \mathbf{I})\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} + [(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1} - (\mathbf{X}^T\mathbf{X})^{-1}]\mathbf{X}^T\mathbf{W}_D\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{W}_D - \mathbf{I})\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} + [(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1} - (\mathbf{X}^T\mathbf{X})^{-1}]\mathbf{X}^T(\mathbf{W}_D - \mathbf{I})\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&\quad + [(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1} - (\mathbf{X}^T\mathbf{X})^{-1}]\mathbf{X}^T\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&= \sigma_n(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} + \sigma_n[(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1} - (\mathbf{X}^T\mathbf{X})^{-1}]\mathbf{X}^T\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&\quad + [(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1} - (\mathbf{X}^T\mathbf{X})^{-1}]\mathbf{X}^T\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&= \sigma_n C_n I_{\mathcal{W}\cap\mathcal{A}} + \sigma_n T_{1n} I_{\mathcal{W}\cap\mathcal{A}} + T_{2n} I_{\mathcal{W}\cap\mathcal{A}} \text{ (say)},
\end{aligned}$$

and thus

$$\sigma_n^{-1}(\hat\beta_{\mathrm{B}} - \hat\beta) = C_n I_{\mathcal{W}\cap\mathcal{A}} + T_{1n} I_{\mathcal{W}\cap\mathcal{A}} + \sigma_n^{-1} T_{2n} I_{\mathcal{W}\cap\mathcal{A}}.$$

Recall that $\mathrm{V_{UBS}} = \sigma_n^{-2}\mathrm{E_B}(\hat\beta_{\mathrm{B}} - \hat\beta)(\hat\beta_{\mathrm{B}} - \hat\beta)^T$. We show that the contributing term in the representation of the bootstrap variance estimate comes from C_n, and the other terms are negligible. We need to compute $\xi^T\mathrm{V_{UBS}}\xi$ for all $\{\xi \in \mathbb{R}^p : ||\xi|| = 1\}$. But since $\mathbf{Q}$ is an orthogonal matrix, we may as well write $\xi = \mathbf{Q}c$, and varying ξ over the unit sphere in $\mathbb{R}^p$ is equivalent to taking all $\{c \in \mathbb{R}^p : ||c|| = 1\}$. We show the following:

$$I_{\mathcal{A}}\mathrm{E_B}\xi^T C_n C_n^T \xi = O_P(pn^{-1}), \tag{5.7}$$

$$I_{\mathcal{A}}\mathrm{E_B}\xi^T C_n C_n^T \xi I_{\mathcal{W}^C} = O_P(p^2n^{-2}), \tag{5.8}$$

$$I_{\mathcal{A}}\mathrm{E_B}\xi^T T_{1n} T_{1n}^T \xi I_{\mathcal{W}} = O_P(\sigma_n^2 p^2 n^{-2}), \tag{5.9}$$

$$I_{\mathcal{A}}\sigma_n^{-2}\mathrm{E_B}\xi^T T_{2n} T_{2n}^T \xi I_{\mathcal{W}} = O_P(p^2n^{-2}), \tag{5.10}$$

$$\sigma_n^{-1} I_{\mathcal{A}}\mathrm{E_B}\xi^T C_n T_{2n}^T \xi I_{\mathcal{W}} = O_P(p^2n^{-2}), \tag{5.11}$$

$$I_{\mathcal{A}}\mathrm{E_B}\xi^T C_n T_{1n}^T \xi I_{\mathcal{W}} = O_P(p^2n^{-2}). \tag{5.12}$$

The first term (5.7) is actually $I_{\mathcal{A}}[\xi^T(\hat{\beta}-\beta)]^2 + O_P(p^2n^{-2})$. For Model 1 since $\mathcal{A}$ is the entire sample space, this implies

$$\mathrm{E_B}\xi^T C_n C_n^T \xi - \xi^T \mathrm{V_n}\xi = O_P(pn^{-3/2}) \tag{5.13}$$

is the leading term, with all other terms negligible. We now proceed to check (5.7)-(5.13).

Proof of (5.7) and (5.13). With an elementary simplification using the singular value decomposition of $\mathbf{X}$, we have

$$\mathrm{E_B}\xi^T C_n C_n^T \xi = \mathrm{E_B} n^{-1} c^T \Lambda^{-1} \mathbf{P}^T \mathbf{W}\mathbf{e}\mathbf{e}^T \mathbf{W}\mathbf{P}\Lambda^{-1} c. \tag{5.14}$$

Recall that $\mathbf{T}_n = diag(e_i^2)$. Now define the following: $\mathbf{M} = n^{-1}c^T\Lambda^{1/2}\mathbf{P}^T\mathbf{T}_n\mathbf{P}\Lambda^{1/2}c$, $\mathbf{M}_1 = c_{11}n^{-1}c^T\Lambda^{1/2}\mathbf{P}^T\mathbf{e}\mathbf{e}^T\mathbf{P}\Lambda^{1/2}c$, $\mathbf{M}_2 = c_{11}n^{-1}c^T\Lambda^{1/2}\mathbf{P}^T\mathbf{T}_n\mathbf{P}\Lambda^{1/2}c$. From (5.14) it can be easily seen that $\mathrm{E_B}\xi^T C_n C_n^T \xi = \mathbf{M} + \mathbf{M}_1 - \mathbf{M}_2$. In proving the results we first note that $\mathbf{P_x} = \mathbf{P}\mathbf{P}^T$. If we define $\eta_b = \sum_{a=1}^p c_a \mathbf{P}_{ba}\lambda_a^{-1/2}$, where the ath element of the vector c is c_a and the ath diagonal entry in the diagonal matrix Λ is λ_a, then on $\mathcal{A}$ we have $\eta_b^2 \le kh_{\max}$, where $h_{\max}$ is the maximum diagonal entry of $\mathbf{P_x}$, known to be $O(p/n)$ from (5.1), and k is some positive constant. Now note that $\mathbf{M} = n^{-1}\sum_{b=1}^n e_b^2\eta_b^2$, $\mathbf{M}_1 = c_{11}n^{-1}(\sum_{b=1}^n e_b\eta_b)^2$, $\mathbf{M}_2 = c_{11}n^{-1}\sum_{b=1}^n e_b^2\eta_b^2$. Hence we can conclude $I_{\mathcal{A}}\mathbf{M}_1 = O_P(p^2n^{-2})$ and $I_{\mathcal{A}}\mathbf{M}_2 = O_P(pn^{-2})$ after some algebra involving (2.5), (1.5), (1.4), (5.1) and (5.2). Similarly, $I_{\mathcal{A}}\mathbf{M}^2 = O_P(p^2n^{-2})$, and note also that $E\mathbf{M} = \xi^T\mathrm{V_n}\xi$. This proves (5.7). For (5.13) we have under Model 1, $\mathbf{M} - \xi^T\mathrm{V_n}\xi = n^{-1}\sum_{b=1}^n(e_b^2 - \tau_b^2)\eta_b^2 = O_P(pn^{-3/2})$.

Proof of (5.8). We only need to look at $I_{\mathcal{A}}\mathrm{E_B}\mathbf{M}I_{\mathcal{W}^C}$, since the other terms are of smaller order. Now $I_{\mathcal{A}}\mathrm{E_B}\mathbf{M}I_{\mathcal{W}^C} = [I_{\mathcal{A}}\mathrm{E_B}\mathbf{M}^2]^{1/2}[1-\mathrm{P_B}(\mathcal{W})]^{\infty/\in} = O_P(p^2n^{-2})$ follows immediately from the proof of (5.7) and Assumption (2.4).

Proof of (5.9). Note that $T_{1n} = -\sigma_n(\mathbf{X}^T\mathbf{W}_D\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{e}$, so $\xi^T T_{1n} I_{\mathcal{W}\cap\mathcal{A}} = -\sigma_n n^{-1/2}c^T\Lambda^{-1/2}\mathbf{U}_B^{-1}\mathbf{P}^T\mathbf{W}\mathbf{P}\mathbf{P}^T\mathbf{W}\mathbf{e}$ and thus

$$\begin{aligned}
&\xi^T T_{1n} T_{1n}^T \xi I_{\mathcal{W}\cap\mathcal{A}} && (5.15)\\
&= \sigma_n^2 n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{U}_B^{-1}\Lambda^{-1/2}cc^T\Lambda^{-1/2}\mathbf{U}_B^{-1}\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} && (5.16)\\
&\le \sigma_n^2 n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{U}_B^{-1}\Lambda^{-1}\mathbf{U}_B^{-1}\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} && (5.17)\\
&\le k\sigma_n^2 n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{U}_B^{-2}\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} && (5.18)\\
&\le \sigma_n^2 n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} && (5.19)\\
&\le \sigma_n^2 n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{A}}. && (5.20)
\end{aligned}$$

The steps involved in the above reduction are as follows: (5.17) follows from (5.16) since $cc^T < \mathbf{I}$; (5.18) follows from (5.17) since $\Lambda^{-1} < k\mathbf{I}$ on $\mathcal{A}$ from condition

(1.5); (5.19) follows from (5.18) since, from Lemma 5.1, $\mathbf{U}_B^{-1} < k\mathbf{I}$ on the set $\mathcal{W} \cap \mathcal{A}$. The last inequality is obvious. Now routine but lengthy algebra shows that the bootstrap expectation of the expression in (5.20) is $O_P(\sigma_n^2 p^2 n^{-2})$.

Proof of (5.10). The proof of this part is similar to that of (5.9). We have $\sigma_n^{-1} T_{2n} = -(\mathbf{X}^T \mathbf{W}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}$, and hence $\sigma_n^{-1} \xi^T T_{2n} I_{\mathcal{W}\cap\mathcal{A}} = -n^{-1/2} c^T \Lambda^{-1/2} \mathbf{U}_B^{-1} \mathbf{P}^T \mathbf{W} \mathbf{P} \mathbf{P}^T \mathbf{e} I_{\mathcal{W}\cap\mathcal{A}}$ and thus

$$\sigma_n^{-2} \xi^T T_{2n} T_{2n}^T \xi I_{\mathcal{W}\cap\mathcal{A}} \leq n^{-1} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P_x} \mathbf{W} \mathbf{P_x} \mathbf{e} I_{\mathcal{A}} \tag{5.21}$$

by following similar steps as in the previous part of the proof. Now again routine algebra shows that the bootstrap expectation of the expression in (5.21) is $O_P(p^2 n^{-2})$.

Proof of (5.11). We have to show $\sigma_n^{-1} \mathrm{E_B} \xi^T C_n T_{2n}^T \xi I_{\mathcal{W}\cap\mathcal{A}} = O_P(p^2 n^{-2})$. Using $\sigma_n^{-1} \xi^T T_{2n} I_{\mathcal{W}\cap\mathcal{A}} = -n^{-1/2} c^T \Lambda^{-1/2} \mathbf{U}_B^{-1} \mathbf{P}^T \mathbf{W} \mathbf{P} \mathbf{P}^T \mathbf{e} I_{\mathcal{W}\cap\mathcal{A}}$, and $\xi^T C_n I_{\mathcal{W}\cap\mathcal{A}} = n^{-1/2} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} I_{\mathcal{W}\cap\mathcal{A}}$, we have $\sigma_n^{-1} \xi^T C_n T_{2n}^T \xi I_{\mathcal{W}\cap\mathcal{A}} = n^{-1} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \mathbf{U}_B^{-1} \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}} = n^{-1} c^T \Lambda^{1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}} + \sigma_n n^{-1} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \mathbf{R}_W \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}$, where Lemma 5.2. states the nature of the matrix $\mathbf{R}_W$. Thus we have

$$\begin{aligned} &\mathrm{E_B} \sigma_n^{-1} \xi^T C_n T_{2n}^T \xi I_{\mathcal{W}\cap\mathcal{A}} \\ &\leq I_{\mathcal{A}} \mathrm{E_B} n^{-1} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \Lambda^{-1/2} c \qquad (5.22) \\ &\quad + \sigma_n n^{-1} \mathrm{E_B} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \mathbf{R}_W \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}. \qquad (5.23) \end{aligned}$$

Consider (5.22) first:

$$\begin{aligned} \mathrm{E_B} n^{-1} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \Lambda^{-1/2} c &\leq k n^{-1} \mathrm{E_B} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P_x} \mathbf{W} \mathbf{e} \\ &= \mathbf{M}_3 + \mathbf{M}_4, \end{aligned} \tag{5.24}$$

where $\mathbf{M}_3 = (1 - c_{11}) k n^{-1} \mathbf{e}^T \mathbf{P_x} \mathbf{M}_5 \mathbf{e}$, $\mathbf{M}_4 = c_{11} k n^{-1} \mathbf{e}^T \mathbf{P_x} \mathbf{e}$, $\mathbf{M}_5 = diag((\mathbf{P_x})_{ii})$. The reduction to (5.24) follows from techniques similar to those used earlier, for example in case of (5.21). Now it is easily verified that $I_{\mathcal{A}} \mathbf{M}_4$ is $O_P(pn^{-2})$. A direct computation using (1.8)-(1.10) shows that $I_{\mathcal{A}} \mathbf{M}_3 = O_P(p^2 n^{-2})$.

Now consider (5.23):

$$\begin{aligned} &\sigma_n n^{-1} |\mathrm{E_B} c^T \Lambda^{-1/2} \mathbf{P}^T \mathbf{W} \mathbf{e} \mathbf{e}^T \mathbf{P_x} \mathbf{W} \mathbf{P} \mathbf{R}_W \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}| \\ &\leq \sigma_n n^{-1} [\mathrm{E_B} ||\mathbf{R}_W \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}||^2]^{1/2} [\mathrm{E_B} ||\mathbf{P}^T \mathbf{W} \mathbf{P_x} \mathbf{e} \mathbf{e}^T \mathbf{W} \mathbf{P} \Lambda^{1/2} c I_{\mathcal{W}\cap\mathcal{A}}||^2]^{1/2}. \end{aligned} \tag{5.25}$$

Now $\mathrm{E_B} ||\mathbf{R}_W \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}||^2 = c^T \Lambda^{-1/2} \mathrm{E_B} \mathbf{R}_W^2 I_{\mathcal{W}\cap\mathcal{A}} \Lambda^{-1/2} c$
$\leq \lambda_{\max}(\mathrm{E_B} \mathbf{R}_W^2 I_{\mathcal{W}\cap\mathcal{A}}) ||\Lambda^{-1/2} c I_{\mathcal{A}}||^2 = O(p^2 n^{-1})$ and

$$\mathrm{E_B} ||\mathbf{P}^T \mathbf{W} \mathbf{P_x} \mathbf{e} \mathbf{e}^T \mathbf{W} \mathbf{P} \Lambda^{-1/2} c I_{\mathcal{W}\cap\mathcal{A}}||^2$$

$$
\begin{aligned}
&= \mathrm{E_B}||\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P}\Lambda^{-1/2}cI_{\mathcal{A}}||^2 \\
&= \mathrm{E_B}[\mathbf{e}^T\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{e}I_{\mathcal{A}}][c^T\Lambda^{-1/2}\mathbf{P}^T\mathbf{W}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P}\Lambda^{-1/2}cI_{\mathcal{A}}] \\
&\le [\mathrm{E_B}(\mathbf{e}^T\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{e}I_{\mathcal{A}})^2\mathrm{E_B}(c^T\Lambda^{-1/2}\mathbf{P}^T\mathbf{W}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P}\Lambda^{-1/2}cI_{\mathcal{A}})^2]^{1/2} \\
&= O_P(p^{3/2}n^{-3/2}) = O_P(p^2n^{-1}).
\end{aligned}
$$

The rates follow from routine calculations. This ensures that (5.25) is $O_P(p^2n^{-2})$.

Proof of (5.12). We have to show $\mathrm{E_B}\xi^TC_nT_{1n}^T\xi I_{\mathcal{W}\cap\mathcal{A}} = O_P(p^2n^{-2})$. As in the previous part of the proof, $\mathrm{E_B}\sigma_n^{-1}\xi^TC_nT_{1n}^T\xi I_{\mathcal{W}\cap\mathcal{A}}$

$$
\le \sigma_n\mathrm{E_B}n^{-1}c^T\Lambda^{-1/2}\mathbf{P}^T\mathbf{W}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\Lambda^{-1/2}cI_{\mathcal{A}} \tag{5.26}
$$

$$
+\sigma_n^2n^{-1}\mathrm{E_B}c^T\Lambda^{-1/2}\mathbf{P}^T\mathbf{W}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{R}_W\Lambda^{-1/2}cI_{\mathcal{W}\cap\mathcal{A}}. \tag{5.27}
$$

Consider (5.26) first: $I_{\mathcal{A}}\mathrm{E_B}n^{-1}c^T\Lambda^{-1/2}\mathbf{P}^T\mathbf{W}\mathbf{e}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\Lambda^{-1/2}c$ $\le kn^{-1}I_{\mathcal{A}}\mathrm{E_B}\mathbf{e}^T\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e} = O_P(p^2n^{-2})$ by direct computation. For (5.27), we have the following:

$$
\begin{aligned}
&||\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{R}_W\Lambda^{-1/2}c||^2I_{\mathcal{W}\cap\mathcal{A}} \\
&= \mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{R}_W^{-1}\Lambda^{-1/2}cc^T\Lambda^{-1/2}\mathbf{R}_W^{-1}\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&\le k\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P}\mathbf{R}_W^{-2}\mathbf{P}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{W}\cap\mathcal{A}} \\
&\le kp^2n^{-1}\mathbf{e}^T\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{P_x}\mathbf{W}\mathbf{e}I_{\mathcal{A}} \text{ (by Lemma 5.2)}
\end{aligned}
$$

and this last is $O_P(p^4n^{-2})$ from calculations as in (5.20). Therefore the quantity in (5.27) is $O_P(\sigma_n^2p^2n^{-2})$. This completes the proof of the theorem.

Proof of Proposition 3.1. We denote m_n by m and note that $Prob[$ at least m of the X's are positive $] = 1-\sum_{i=0}^{m-1} Prob[$ exactly i of the X's are positive $]$. We can exclude the case $i = 0$, since at least one of the X's is always positive. For $j = 1, \ldots, m-1$,

$$
\begin{aligned}
&Prob[\text{ exactly } j \text{ of the } X\text{'s are positive}] \\
&= \frac{n!}{n^n}\binom{n}{j}\sum_{a_1+a_2+\cdots+a_j=n-j, a_i\ge 0}\frac{1}{(a_1+1)!(a_2+1)!\ldots(a_j+1)!} \\
&\le \frac{n!j^{(n-j)}}{n^n(n-j+1)(n-j)!}\binom{n}{j}\sum_{a_1+a_2+\cdots+a_j=n-j, a_i\ge 0}\frac{(n-j)!}{(a_1)!(a_2)!\ldots(a_j)!}\frac{1}{j^{n-j}} \\
&= \frac{n!n!j^{(n-j)}}{n^n(n-j+1)!j!(n-j)!}.
\end{aligned}
$$

The inequality in the middle comes from $(a_1+1)(a_2+1)\ldots(a_j+1) \ge 1+\sum_{i=1}^j a_i$, since all the a_i's are non-negative, and then we use $\sum_{i=1}^j a_i = n-j$.

We can bound $\frac{n!n!j^{(n-j)}}{n^n(n-j+1)!j!(n-j)!}$ above by $n^{-1}(\frac{j}{n})^{n-2j}\frac{j^j}{j!}$, and using an upper bound in the Stirling's approximation of $\frac{j^j}{j!}$ from Feller (1968, p.54), we see that a further upper bound is $n^{-1}(\frac{j}{n})^{n-2j}\frac{e^j}{\sqrt{2\pi j}}$, and across $j = 1, \ldots, m-1$, this is bounded above by $n^{-1}(\frac{m}{n})^{n-2m}\frac{e^m}{\sqrt{2\pi}}$. Now

$$\sum_{i=0}^{m-1} Prob[\text{ exactly } i \text{ of the } X\text{'s are positive}] \leq \sum_{i=0}^{m-1} n^{-1}(\frac{m}{n})^{n-2m}\frac{e^m}{\sqrt{2\pi}}$$
$$\leq (\frac{m}{n})^{n-2m}\frac{e^m}{\sqrt{2\pi}}$$
$$\leq \frac{1}{\sqrt{2\pi}} c_2^{n-2c_2n} e^{c_2n} \text{ (since } m \leq c_2n \leq n/2)$$
$$= \frac{1}{\sqrt{2\pi}} e^{-n[(1-2c_2)\Delta - c_2]} \text{ (putting } -\log(c_2) = \Delta).$$

We only have to show that $\alpha = -(1-2c_2)\log(c_2) - c_2 > 0$, which is true if c_2 is small enough. An elementary computation shows that $\alpha = 0$ near $c_2 = 0.34128$, and taking $c_2 = 1/3$ yields $\alpha > 0$.

Acknowledgement

We thank the anonymous referee for his detailed, constructive and thoughtful comments and suggestions. This has lead to a significant improvement of this paper. We also thank Mr. J. P. N. Bishwal for drawing our attention to Wick's lemma.

References

Barbe, P. and Bertail, P. (1995). *The Weighted Bootstrap.* Lecture Notes in Statistics **98**. Springer - Verlag, New York.

Bera, A. K. and Higgins, M. L. (1993). A survey of ARCH models: properties, estimation and testing. *J. Econom. Surveys* **7**, 305-366.

Bickel, P. J. and Freedman, D. A. (1983). Bootstrapping regression models with many parameters. In *Festschrift for Erich L. Lehmann* (Edited by P. Bickel, K. Doksum and J. Hodges, Jr.), 28-48. Wadsworth, Belmont, Calif.

Boos, D. and Monahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika* **73**, 77-83.

Chatterjee, S. (1998). Another look at the jackknife: further examples of the generalised bootstrap. *Statist. Probab. Lett.* **40**, 307-319.

Chatterjee, S. and Bose, A. (1998). Dimension asymptotics for the uncorrelated weights bootstrap in linear regression. *Tech. Rep.* **12/98**, Stat-Math. Unit, Indian Statistical Institute.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 101-118.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Volume-1. 3rd edition. Wiley Eastern, New Delhi.

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218-1228.

Härdle, W. and Marron, S. (1991). Bootstrap simultaneous error bars for non-parametric regression. *Ann. Statist.* **19**, 778-796.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19**, 285-292.

Liu, R. Y. and Singh, K. (1992). Efficiency and robustness in resampling. *Ann. Statist.* **20**, 370-384.

Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* **15**, 360-375.

Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā, Ser. A* **53**, 320-333.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255-285.

Praestgaard, J. and Wellner, J. A. (1993). Exchangeable weighted bootstrap of the general empirical process. *Ann. Probab.* **21**, 2053-2086.

Rubin, D. (1981). The Bayesian bootstrap. *Ann. Statist.* **9**, 130-134.

Shao, J. and Wu, C. F. J. (1987). Heteroskedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15**, 1563-1579.

Weng, C. S. (1989). On a second order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17**, 705-710.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1295.

Zheng and Tu (1988). Random weighting method in regression models. *Sci. Sinica Ser. A.* **31**, 1442-1459.

Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta 700035, India.

E-mail: abose@isical.ac.in

E-mail: res9514@isical.ac.in

(Received November 1997; accepted August 1999)