Ann. Inst. Statist. Math.
Vol. 54, No. 2, 367–381 (2002)

DIMENSION ASYMPTOTICS FOR GENERALISED BOOTSTRAP IN LINEAR REGRESSION

SNIGDHANSU CHATTERJEE[1]* AND ARUP BOSE[2]

[1]*Department of Mathematics and Statistics, University of Nebraska-Lincoln, 924 Oldfather Hall, P.O. Box 880323, Lincoln, NE 68588-0323, U.S.A.,* e-mail: schatterjee@math.unl.edu
[2]*Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India,* e-mail: abose@isical.ac.in

(Received May 17, 1999; revised May 29, 2000)

Abstract. We prove consistency of a class of generalised bootstrap techniques for the distribution of the least squares parameter estimator in linear regression, when the number of parameters tend to infinity with data size and the regressors are random. We show that best results are obtainable with resampling techniques that have not been considered earlier in the literature.

Key words and phrases: Bootstrap, jackknife, regression, dimension asymptotics.

1. Introduction

Consider the linear model

$$y_{i:n} = \boldsymbol{x}_{i:n}^T \beta_n + e_{i:n}, \qquad i = 1, \dots, n \tag{1.1}$$

where $y_{i:n}$ are n observations, $\boldsymbol{x}_{i:n}$ are the observed values of random design vectors in $\mathbb{R}^p$, β_n is a p dimensional unknown but constant parameter vector and $e_{i:n}$ are unobservable noise terms. The dimension of the parameter, p is allowed to tend to infinity with data size. In many applications, models are used where the dimension p is not small compared to n. Then an asymptotic approach where p is fixed is misleading since the high dimensionality of the model is lost asymptotically. Asymptotics where p may increase with n has been considered in Huber (1981), Haberman (1977*a*, 1977*b*), Shorack (1982), Bickel and Freedman (1983), Portnoy (1984, 1985, 1988), Mammen (1989) and Sauermann (1989) in different contexts. Mammen (1989) has appropriately named this kind of asymptotic study *dimension asymptotics*.

Let $\hat{\beta}$ be the least squares estimator of β. For a vector $c = (c_1 \cdots c_n)^T$, consider the problem of estimating the distribution of $n^{1/2} c^T (\hat{\beta} - \beta)$ as $n \to \infty$. Mammen (1993) has shown that paired bootstrap and wild bootstrap methods provide consistent estimators. In this paper we extend the paired bootstrap based consistency result of Mammen (1993) to the case of resampling with random weights, which is often termed *generalised bootstrap* in the literature.

Bootstrap with random weights was probably considered for the first time in Rubin (1981). Several other authors have considered special cases of random weights bootstrap,

*This research was carried out when with the Theoretical Statistics and Mathematics Unit, Indian Statistical Institute.

for example see Zheng and Tu (1988), Lo (1991). A detailed review of generalised bootstrap may be found in Barbe and Bertail (1995). In the context of regression, however, the above references considered only generalised residual based bootstrap, and not the paired bootstrap. Generalisations of paired bootstrap and delete-d jackknives are available in Chatterjee and Bose (2000a) and Chatterjee (1998). In this paper we concentrate on generalisation of the paired bootstrap.

Our choice of generalised paired bootstrap over residual based resampling techniques was guided by an important robustness consideration. Mammen (1996) showed that in high dimensional regression models, if the estimator is obtained under wrong assumptions, the empirical process of residuals tends to be biased towards the wrong assumption. This is true even if the linear fit from the 'wrong' estimator is consistent, see ($A5$) in Mammen (1996). Loosely speaking, this reflects the phenomenon that the residuals reflect the assumptions and not the nature of the data. Since the residuals are sensitive to the assumptions, bootstrap based on residuals is also expected to exhibit such sensitivity.

Our conditions on weights are similar to those used in resampling theory based on empirical processes. However, there are significant differences between the general empirical process based approach that is widely used and our approach to resampling in the present paper. Suppose, for some statistic, P, P_n, P_{Bn} are the actual law, the sample (empirical) law and the bootstrap law, respectively. The major thrust of empirical process literature is usually to prove both

$$n^{1/2}(P_n - P) \Rightarrow G_P$$
$$n^{1/2}(P_{Bn} - P_n) \Rightarrow G_P$$

either almost surely or in probability, where G_P is a P-Brownian bridge. This implies usual bootstrap consistency, that is, $F_{Bn}(t) - F_n(t) \to 0$ uniformly over t, where F denotes the distribution function. For bootstrapping with random weights, usual assumptions on the weights are that they are either independent or exchangeable, with some restrictions on their lower order moments. Let $W_{i:n}$ be the weights used for resampling. Then, along with other conditions, Praestgaard and Wellner (1993) assume that

$$n^{-1}\sum_{i=1}^{n}(W_{i:n}-1)^2 \xrightarrow{P} c^2 > 0. \tag{1.2}$$

See also Mason and Newton (1992) , where an equivalent of (1.2) is given in $\mathcal{W}_{III}$. The assumption (1.2) is only slightly weaker than $\sigma_n^2 = \mathrm{Var}(W_{i:n}) \to c^2 > 0$.

A major consideration in dimension asymptotics is to obtain the highest possible rate of growth of p with respect to n. The paired bootstrap typically requires $p^4/n^3 \to 0$ (Mammen (1993)). Let $\{w_{i:n}\}$ be the sequence of non-negative, exchangeable weights used for resampling, and let $\sigma_n^2 = \mathrm{Var}(w_{i:n})$. Our major result in this paper is to show that by letting $\sigma_n^2 \to 0$, we can have best dimension asymptotic results. That is, distributional consistency results hold for generalised bootstrap even under the stringent condition $p/n \to 0$ when σ_n^2 tends to zero at an appropriate rate. The rate $p/n \to 0$ is the best achievable when bootstrapping from residuals, as proved by Bickel and Freedman (1983). We also study consistency of resampling with random weights whose variance satisfies a positive lower bound, for example the delete-d jackknives with $d/n \to c \in (0,1)$ and the paired bootstrap. For these techniques, we show consistency when $p^2/n \to 0$.

The consistency of m out of n bootstrap is also established, when $m \to \infty$ and $m/n \to 0$. Depending on the dimension p, a criterion for the choice of m is also suggested. Our model set-up, discussed in details in Section 3, is same as that of Mammen (1993).

It has been established in Mammen (1993) that the paired bootstrap is consistent even when parameter dimension is increasing, regressors are random and errors display heteroscedasticity. Similar robustness results have been established in Chatterjee and Bose (2000a) for generalised paired bootstrap in the context of variance estimation. However, resampling schemes based on data pairs are not expected to be second order accurate for distribution approximation. They do not capture the first term in the Edgeworth expansion of the distribution of the least squares estimator. See Wu (1990), Mammen (1993) and Hall and Mammen (1994) for some discussion on this. However, after a bias correction, higher order accuracy results are obtainable for some schemes for which the variance of weights remains bounded away from zero (Chatterjee and Bose (2000b)). It is not known whether higher order accuracy is obtainable for asymptotically degenerate weights for which in this paper we get the best dimension asymptotic results.

Using asymptotically degenerate weights essentially means that the bootstrap empirical distribution function "cannot be too far" from the original empirical distribution function, in case the parameter dimension is large. This heuristic interpretation of our result may be useful in resampling for infinite dimensional parameters also, such as in nonparametric or semiparametric inference.

2. The resampling technique

We will henceforth drop the suffix n from the above notations, thus $y_i = y_{i:n}$, $\boldsymbol{x}_i = \boldsymbol{x}_{i:n}$, $\beta = \beta_n$ and $e_i = e_{i:n}$ from now on. We will also have occasion to use the usual matrix notation for the above, thus $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T$, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^T$ and $\boldsymbol{e} = (e_1, e_2, \ldots, e_n)^T$.

In this section we discuss the resampling scheme proposed by Chatterjee and Bose (2000a) and Chatterjee (1998) which we use here. In regression problems, the paired bootstrap proceeds by taking an $SRSWR$ sample of size n from the data pairs $\{(\boldsymbol{x}_i, y_i), i = 1, \ldots, n\}$. This is same as transforming the data to $\{\sqrt{w_i}(\boldsymbol{x}_i, y_i), i = 1, \ldots, n\}$ where $(w_1, \ldots, w_n)$ come from *Multinomial* $(n, 1/n, \ldots, 1/n)$. Instead of multinomial weights, the generalised bootstrap is carried out by using any non-negative exchangeable random weights satisfying the conditions that we mention now. Let $V(w_i) = \sigma_n^2$. We adopt the notation

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

Also 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 $\geq 1/3$.

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

$$0 < \sigma_n^2 = o(np^{-5/2}) \tag{2.2}$$

$$P_B[\mathcal{W}] = 1 - O(pn^{-1}) \tag{2.3}$$

$$c_{11} = O(n^{-1}) \tag{2.4}$$

$$\forall i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 3, \qquad c_{i_1 i_2 \cdots i_k} = O(n^{-k+1}) \tag{2.5}$$

$$(2.6) \qquad \forall i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 4, \qquad c_{i_1 i_2 \cdots i_k} = O(\min(n^{-k+2}, 1)).$$

The bootstrap estimator is defined as

$$\hat{\beta}_B = \left(\sum w_i \boldsymbol{x}_i \boldsymbol{x}_i^T\right)^{-1} \sum w_i \boldsymbol{x}_i y_i I_{\mathcal{W}} + \hat{\beta}(1 - I_{\mathcal{W}}).$$

Condition (2.3) ensures that $\hat{\beta}_B$ is the first term on the right hand side of the above relation a large number of times. Thus $\hat{\beta}_B$ is realised by ordinary least squares on the randomly weighted data $\{\sqrt{w_i}(\boldsymbol{x}_i, y_i), i = 1, \ldots, n\}$, with a correction for the cases when a substantial number of weights are zero.

Several known resampling techniques are special cases of the present set-up. Among them are the paired bootstrap, the delete-d jackknife, the m out of n bootstrap where $m \to \infty$ and $m/n \to 0$; and several choices of weights used in Bayesian bootstrap schemes. Another choice is i.i.d. weights whose support satisfies a positive lower bound, and whose mean and variance match with the above restrictions; the other conditions being trivially satisfied. Such weights have the advantage of being convenient for practical use, being easily programmable and requiring less computer time and memory.

3. Main result

We assume

$$(3.1) \qquad \text{For every } n, (\boldsymbol{x}_i, y_i) 1 \le i \le n \text{ are i.i.d.}$$

$$(3.2) \qquad \sup_n E y_{1:n}^2 < \infty,$$

$$(3.3) \qquad \sup_n E\|\boldsymbol{x}_{1:n}\|^2 < \infty.$$

We also assume

$$(3.4) \qquad \beta \text{ minimises } b \text{ in } E(y_i - \boldsymbol{x}_i^t b)^2.$$

A careful study of the results of this paper and Mammen (1993) reveals that in (3.1) the independence of $(\boldsymbol{x}_i, y_i)$ across i is usually used, and the assumption about identical nature of their distribution may be dropped in favour of some simple moment conditions. However, we retain the i.i.d. assumption for simplicity.

We assume that $E\boldsymbol{x}_i\boldsymbol{x}_i^T$ is nonsingular, so that β is uniquely defined. With the additional assumption $E(y_i - \boldsymbol{x}_i^T\beta)^2 > 0$, Mammen (1993) has discussed that the following are always fulfilled after a standardisation:

$$(3.5) \qquad E\boldsymbol{x}_i\boldsymbol{x}_i^T = \boldsymbol{I}, \qquad i = 1, \ldots, n$$

$$(3.6) \qquad 0 < \inf_n E e_i^2 \le \sup E e_i^2 < \infty, \qquad i = 1, \ldots, n.$$

As Mammen (1993) points out, the model conditions imply $E\boldsymbol{x}_i e_i = 0, \forall i$, but they do not imply the traditional homoscedastic assumption that given $\boldsymbol{X}$ the e_i are conditionally i.i.d. with mean zero. Also, the parameter β actually defines the linear model 'nearest' to the data, which is often a useful thing to consider even when linear models do not hold exactly.

We denote $\mathcal{L}(\cdot \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, y_1, \ldots, y_n)$ by $\mathcal{L}^B$. The sup-norm distance between distributions is denoted by d_∞. The notations used in this paper follow Mammen (1993) as far as feasible, while in the context of generalised bootstrap sometimes notations of Section 2 are also used.

THEOREM 3.1. *Choose $c \in I\!R^p$ with $\|c\| = 1$. Consider a linear model (1.1) satisfying (3.1)–(3.6). For a fixed δ, set K to be the smallest integer greater than or equal to $2/\delta$ and assume that*

(3.7) $$p^{1+\delta}/n \to 0$$

(3.8) $$\sup_n \sup_{\|d\|=1} E|d^T\boldsymbol{x}_i|^{4K}(1 + e_i^2) < \infty$$

(3.9) $$E|(c^T\boldsymbol{x}_i)e_i|^2 > 0$$

(3.10) $$E|(c^T\boldsymbol{x}_i)e_i|^{2+\gamma} < \infty \quad \text{for some} \quad \gamma > 0.$$

Consider resampling with weights which satisfy $c_{22} \to 1$ in addition to conditions given in Section 2. Then

(3.11) $$d_\infty(\mathcal{L}^B(n^{1/2}{\sigma_n}^{-1}c^T(\hat{\beta}_B - \hat{\beta})), \quad \mathcal{L}(n^{1/2}c^T(\hat{\beta} - \beta))) \to 0 \text{ in probability}$$

holds if either of the following holds:

(3.12) $$\delta \geq 1 \quad \text{and} \quad \sigma_n^2 > k > 0\,\forall n$$

(3.13) $$1 > \delta \geq 1/3 \quad \text{and} \quad \sigma_n^2 = O(n^{-1})\,\forall n$$

(3.14) $$1 > \delta \geq 0 \quad \text{and} \quad \sigma_n^4 p^5 n^{-2} \to 0.$$

Remark 1. (a) Condition (3.14) ensures that $\delta = 0$ (i.e. $p/n \to 0$) is enough if $\sigma_n^2 = o(n^{-3/2})$.

(b) For the m out of n bootstrap with replacement, with $m \to \infty$ and $m/n \to 0$, we have $\sigma_n^2 \approx n/m$, consistency is achieved if either $p^2/n \to 0$ or if $\delta < 1$ and $p^5/m^2 \to 0$. Note that this restricts the choice of m to require that $nm^{-2(1+\delta)/5} \to 0$. This is a criterion for choice of m.

Remark 2. Theorem 3.1 implies that the paired bootstrap is consistent if $p^2/n \to 0$, which is weaker than the rate $p^4/n^3 \to 0$ obtained by Mammen (1993). This is because in our general framework for resampling, we do not use the special properties of the paired bootstrap. Mammen (1993) has also shown that if $E[e_i \mid \boldsymbol{x}_i] = 0$ and $p/n \to 0$, then the paired bootstrap is consistent. For our results, we do not need any condition on conditional expectation of errors. Note that the resampling consistency result (3.11) as well as its corresponding result of Mammen (1993) are convergence in probability results. This is due to the fact that a triangular array of random variables are involved.

Remark 3. We have replaced the Lindeberg condition of Mammen (1993) with a Lyapunov-type condition in (3.10). This is for convenience in the bootstrap distributional asymptotics. Conditions (3.7) and (3.8) are such that weakening of one leads to strengthening of the other. Mammen (1993) had conjectured, following Yin *et al.* (1988), that for the design where elements of $\boldsymbol{x}_i$ are i.i.d., finite fourth moments of the elements of $\boldsymbol{x}_i$ would suffice.

Remark 4. Two important points about the results, that have been pointed out by Mammen (1993). The first is that if $E(e_i \mid \boldsymbol{x}_i) \neq 0$, then the bias of $n^{1/2}c^T(\hat{\beta}-\beta)$ is of the order $p/n^{1/2}$, which may tend to ∞ if $\delta < 1$. Thus in this case for the mean zero normal approximation to work requires $\delta \geq 1$. The other is that the conditional law $\mathcal{L}(n^{1/2}c^T(\hat{\beta}-\beta) \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is also consistently estimated by generalised bootstrap. This is because, as will be seen later, both the conditional and unconditional laws converge to the same limit $N(0, E(c^T\boldsymbol{x}_1)^2 e_1^2)$.

Remark 5. An important example of asymptotically degenerate weights is the delete-d jackknife with $d/n \to 0$. Wu (1990) has shown that these jackknives are generally inconsistent. Note that for these schemes the fourth central moment of the weights is $O(n)$ and hence ruled out by our assumptions.

4. Proofs

In order to prove the theorem, we need some results which we state now. Note that for resampling with generalised bootstrap weights, we have $\boldsymbol{x}_i^* = \sqrt{w_i}\boldsymbol{x}_i$, $y_i^* = \sqrt{w_i}y_i$, $e_i^* = \sqrt{w_i}e_i$. The matrix $\boldsymbol{W}$ is the diagonal matrix with i-th diagonal entry w_i, and $W_i = \sigma_n{}^{-1}(w_i - 1)$. Define $A = \boldsymbol{I} - n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T$ and $A^* = \boldsymbol{I} - n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^*\boldsymbol{x}_i^{T*}$. Also let $\lambda_{\text{amax}}(A)$ be the maximum of the absolute values of the eigenvalues of A. We use the notation $\sum^{\neq}$ to denote sum over indices that are different from one another. Now

LEMMA 4.1. *Under the conditions of Theorem* 3.1

$$E_B\lambda_{\text{amax}}(A^{*4}) = O_P(\max\{p^{2+\delta}n^{-2}, \sigma_n^4 p^3 n^{-2}\}). \tag{4.1}$$

LEMMA 4.2. *Under the assumptions of Theorem* 3.1, *on the set where both* $\lambda_{\text{amax}}(A) < 1$ *and* $\lambda_{\text{amax}}(A^*) < 1$, *we have*

$$n^{1/2}\sigma_n{}^{-1}c^T(\hat{\beta}_B - \hat{\beta}) = n^{-1/2}\sum_{i=1}^n W_i c^T\boldsymbol{x}_i e_i + r_{nB} + o_P(1) \tag{4.2}$$

where $P_B[|r_{nB}| > \epsilon] \to^p 0$ *for any* $\epsilon > 0$.

The proof of the above results are given later. Conditions (3.7), (3.8), (3.12)–(3.14) are required for the proof of the above two lemma.

PROOF OF THEOREM 3.1. One part of the proof involves the arguments to show that $\mathcal{L}(n^{1/2}c^T(\hat{\beta}-\beta)) \to N(0, E(c^T\boldsymbol{x}_1)^2 e_1^2)$, and this is identical with Mammen (1993), so we will not repeat it here. Note that $P[\lambda_{\text{amax}}(A) > 1] = o(1)$ and $P_B[\lambda_{\text{amax}}(A^*) > 1] = o_P(1)$. This is shown later. Thus by using Lemma 4.2 it is sufficient to show that $\mathcal{L}^B(n^{1/2}\sigma_n{}^{-1}c^T(\hat{\beta}_B - \hat{\beta})) \to N(0, E(c^T\boldsymbol{x}_1)^2 e_1^2)$ in probability. Now

$$\begin{aligned}
&P_B[n^{1/2}c^T(\hat{\beta}_B - \hat{\beta}) \leq x] \\
&\quad = P_B[n^{1/2}c^T(\hat{\beta}_B - \hat{\beta}) \leq x, I_{\mathcal{W}} = 1] + P_B[n^{1/2}c^T(\hat{\beta}_B - \hat{\beta}) \leq x, I_{\mathcal{W}} = 0] \\
&\quad \leq P_B[n^{1/2}c^T(\hat{\beta}_B - \hat{\beta}) \leq x, I_{\mathcal{W}} = 1] + P_B[I_{\mathcal{W}} = 0] \\
&\quad = P_B[n^{1/2}c^T(\hat{\beta}_B - \hat{\beta}) \leq x, I_{\mathcal{W}} = 1] + o_P(1).
\end{aligned}$$

For the rest of this paper it is implicitly assumed that calculations are carried out on the set $\mathcal{W}$, although we drop the indicator of this set to simplify notations. Note that

$$\begin{aligned} P_B[n^{1/2}c^T(\hat\beta_B-\hat\beta)\le x] &= P_B[n^{1/2}c^T(\hat\beta_B-\hat\beta)\le x, \lambda_{\text{amax}}(A^*)<1] \\ &\quad + P_B[n^{1/2}c^T(\hat\beta_B-\hat\beta)\le x, \lambda_{\text{amax}}(A^*)\ge 1] \\ &= P_B[n^{1/2}c^T(\hat\beta_B-\hat\beta)\le x, \lambda_{\text{amax}}(A^*)<1]+o_P(1) \end{aligned}$$

since

$$\begin{aligned} P_B[n^{1/2}c^T(\hat\beta_B-\hat\beta)\le x, \lambda_{\text{amax}}(A^*)\ge 1] &\le P_B[\lambda_{\text{amax}}(A^*)\ge 1] \\ &\le E_B[\lambda_{\text{amax}}(A^{*2})] \\ &= o_P(1) \text{ by Lemma 4.1.} \end{aligned}$$

But on the set $\lambda_{\text{amax}}(A^*)<1$, from Lemma 4.2, (4.2) holds. To complete the proof of the theorem we need to establish that

$$\mathcal{L}_B\left(n^{-1/2}\sum_{i=1}^n W_i c^T \boldsymbol{x}_i e_i\right) \to N(0, E(c^T\boldsymbol{x}_1)^2 e_1^2) \tag{4.3}$$

in probability. For this part of the proof, let $\xi_i = c^T\boldsymbol{x}_i e_i$. Note that for the bootstrap distribution, which is conditional on the data, ξ_i's are constants. In order to prove the above central limit theorem, we use Lemma 4.6 of Praestgaard and Wellner (1993). We quote this lemma below:

THEOREM 4.1. (Praestgaard and Wellner (1993)) *Let $\{m\}$ be a sequence of natural numbers, let $\{a_{mj}\}$ be a triangular array of constants, and let B_{mj}, $j=1,\ldots,m$, $m\in\{m\}$ be a triangular array of row-exchangeable random variables such that*

$$m^{-1}\sum_{j=1}^m (a_{mj}-\bar a_m)^2 \to \tau^2>0, \qquad m^{-1}\max_{j=1,\ldots,m}(a_{mj}-\bar a_m)^2\to 0$$

$$m^{-1}\sum_{j=1}^m (B_{mj}-\bar B_m)^2 \xrightarrow{p} c^2>0, \qquad \lim_{K\to\infty}\limsup_{m\to\infty} E(B_{mj}-\bar B_m)^2 I_{\{|B_{mj}-\bar B_m|>K\}}=0.$$

Then

$$\frac{1}{\sqrt{m}}\sum_{j=1}^m (a_{mj}B_{mj}-\bar a_m \bar B_m) \Rightarrow N(0,c^2\tau^2). \tag{4.4}$$

In our set-up, conditional on the data, $m=n$, $a_{mi}=\xi_i$ and $B_{mi}=W_i$. The first two conditions are satisfied in probability from (3.9) and (3.10), with $\tau^2=E\xi_1^2$. Also, since ξ_i and W_i are mean zero random variables, the mean adjustments in (4.4) may be ignored. Note that variance of W_i is identically one, so the third condition of the theorem is satisfied. From Lemma 4.7 and Lemma 3.1 of Praestgaard and Wellner (1993), the last condition of the above theorem is satisfied because of (2.1) and (2.6). This completes the proof of the theorem. □

PROOF OF LEMMA 4.1. Let

$$A_B = A^* - A = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T - n^{-1}\sum_{i=1}^n \boldsymbol{x}_i^*\boldsymbol{x}_i^{T*} = -\sigma_n n^{-1}\sum_{i=1}^n W_i\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

If $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ respectively denote the maximum and minimum eigenvalue, then it can be easily seen

$$\lambda_{\max}(A^*) \leq \lambda_{\max}(A) + \lambda_{\max}(A_B), \qquad \lambda_{\min}(A^*) \geq \lambda_{\min}(A) + \lambda_{\min}(A_B).$$

Hence we have $\lambda_{\text{amax}}(A^{*4}) \leq 8[\lambda_{\text{amax}}(A^4) + \lambda_{\text{amax}}(A_B^4)]$. From Lemma 1 of Mammen (1993),

$$\lambda_{\text{amax}}(A^4) = O_P(p^{2+\delta}n^{-2}). \tag{4.5}$$

Thus we only have to establish the rates for A_B. With some simple algebra involving the moments of the weights and (3.8), it can be shown that $E_B \operatorname{tr}(A_B^4) = O_P(\sigma_n^4 n^{-2} p^3)$. Thus we have

$$E_B \lambda_{\text{amax}}(A_B^4) = O_P(\sigma_n^4 p^3 n^{-2}). \tag{4.6}$$

Putting together the results of (4.5) and (4.6) the lemma is established. □

In order to prove Lemma 4.2 we need two more results which we now state and prove. For any i, define $A_i = \sum_{j\neq i} \boldsymbol{x}_j \boldsymbol{x}_j^T - nI$. Observe that $A = n^{-1}A_i + n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T$, and A_i and $\boldsymbol{x}_i$ are independent. We employ this fact repeatedly in proofs that come later, so it is useful to discuss the rate for $\lambda_{\text{amax}}(A_i)$.

LEMMA 4.3. *Under the conditions of Theorem* 3.1

$$\sup_i E\lambda_{\text{amax}}\left(\frac{A_i}{n}\right)^{2K} = O(p^{K+K\delta/2}n^{-K}).$$

PROOF OF LEMMA 4.3. Note that for any i, $n^{-1}A_i = A - n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T$. We have

$$\lambda_{\text{amax}}\left(\frac{A_i}{n}\right)^{2K} \leq 2^{2K-1}[\lambda_{\text{amax}}(A^{2K}) + \lambda_{\text{amax}}(n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T)^{2K}].$$

Lemma 1 of Mammen (1993) establishes that $E\lambda_{\text{amax}}(A^{2K}) = O(p^{K+K\delta/2}n^{-K})$. Further $n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T$ is a rank 1 matrix, with its only non-zero eigenvalue being $n^{-1}\|\boldsymbol{x}_i\|^2$. Now from Lemma 0 of Mammen (1993), we have $E\|\boldsymbol{x}_i\|^{4K}p^{-2K}$ is uniformly bounded over i, so that $E\lambda_{\text{amax}}(n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T)^{2K}$ is $O(p^{2K}n^{-2K})$ uniformly over i. This concludes the proof. □

Now define $\hat{\beta}^K = \beta + n^{-1}\sum_{h=1}^{K-1} A^h \sum_{i=1}^n \boldsymbol{x}_i e_i$ and $\hat{\beta}_B^K = \beta + n^{-1}\sum_{h=1}^{K-1} A^{*h} \sum_{i=1}^n \boldsymbol{x}_i^* e_i^*$.

LEMMA 4.4. *Under the assumptions of Theorem* 3.1, *on the set where both* $\lambda_{\text{amax}}(A) < 1$ *and* $\lambda_{\text{amax}}(A^*) < 1$, *we have* $n^{1/2}\sigma_n{}^{-1}c^T(\hat{\beta}_B^K - \hat{\beta}^K) = r_{nB}$ *where* $P_B[|r_{nB}| > \epsilon] \to^p 0$ *for any* $\epsilon > 0$.

PROOF OF LEMMA 4.4. Note that

$$(4.7)\quad n^{1/2}{\sigma_n}^{-1}c^T(\hat{\beta}_B^K-\hat{\beta}^K) = n^{-1/2}{\sigma_n}^{-1}c^T\left(\sum_{h=1}^{K-1}A^{*h}\sum_{i=1}^{n}\boldsymbol{x}_i^*e_i^* - \sum_{h=1}^{K-1}A^h\sum_{i=1}^{n}\boldsymbol{x}_ie_i\right)$$

$$= n^{-1/2}{\sigma_n}^{-1}c^T\sum_{h=1}^{K-1}\left[(A^{*h}-A^h)\sum_{i=1}^{n}w_i\boldsymbol{x}_ie_i + A^h\sum_{i=1}^{n}(w_i-1)\boldsymbol{x}_ie_i\right].$$

The sum over h is a finite sum with K being a constant, so it suffices to consider the individual terms in the summations and establish the following three results for every h:

$$(4.8)\qquad E_B\left[n^{-1/2}{\sigma_n}^{-1}c^TA^h\sum_{i=1}^{n}(w_i-1)\boldsymbol{x}_ie_i\right]^2 = o_P(1)$$

$$(4.9)\qquad E_B\left|n^{-1/2}{\sigma_n}^{-1}c^T(A^{*h}-A^h)\sum_{i=1}^{n}(w_i-1)\boldsymbol{x}_ie_i\right| = o_P(1)$$

$$(4.10)\qquad E_B\left|n^{-1/2}{\sigma_n}^{-1}c^T(A^{*h}-A^h)\sum_{i=1}^{n}\boldsymbol{x}_ie_i\right| = o_P(1).$$

PROOF OF (4.8).

$$E_B\left[n^{-1/2}{\sigma_n}^{-1}c^TA^h\sum_{i=1}^{n}(w_i-1)\boldsymbol{x}_ie_i\right]^2$$

$$= n^{-1}E_B\left[\sum_{i=1}^{n}W_i(c^TA^h\boldsymbol{x}_i)e_i\right]^2$$

$$= n^{-1}\sum_{i=1}^{n}(c^TA^h\boldsymbol{x}_i)^2e_i^2 + n^{-1}c_{11}\sum_{ij}^{\neq}(c^TA^h\boldsymbol{x}_i)e_i(c^TA^h\boldsymbol{x}_j)e_j$$

$$(4.11)\qquad = n^{-1}(1-c_{11})\sum_{i=1}^{n}(c^TA^h\boldsymbol{x}_i)^2e_i^2 + n^{-1}c_{11}\left[\sum_{i=1}^{n}(c^TA^h\boldsymbol{x}_i)e_i\right]^2.$$

From Lemma 2 of Mammen (1993) we know that $n^{-1}[\sum_{i=1}^n(c^TA^h\boldsymbol{x}_i)e_i]^2 = O_P(p)$, and since $c_{11}=O(n^{-1})$, the second expression in (4.11) is $O_P(p/n)=o_P(1)$. For the other quantity in (4.11), recall that

$$(4.12)\qquad A = n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{x}_i^T-\boldsymbol{I}) = n^{-1}A_i + n^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^T.$$

Thus we have

$$n^{-1}\sum_{i=1}^{n}(c^TA^h\boldsymbol{x}_i)^2e_i^2 = n^{-3}\sum_{i=1}^{n}[c^T(A_i+\boldsymbol{x}_i\boldsymbol{x}_i^T)A^{h-1}\boldsymbol{x}_i]^2e_i^2$$

$$\leq 2n^{-3}\sum_{i=1}^{n}[(c^T A_i A^{h-1}\boldsymbol{x}_i)^2 e_i^2 + (c^T \boldsymbol{x}_i \boldsymbol{x}_i^T A^{h-1}\boldsymbol{x}_i)^2 e_i^2]. \tag{4.13}$$

Consider the second term in the above expression. We have

$$n^{-3}\sum_{i=1}^{n}[c^T \boldsymbol{x}_i \boldsymbol{x}_i^T A^{h-1}\boldsymbol{x}_i]^2 e_i^2 \leq n^{-3}\lambda_{\text{amax}}(A^{2(h-1)})\sum_{i=1}^{n}[c^T\boldsymbol{x}_i]^2\|\boldsymbol{x}_i\|^4 e_i^2. \tag{4.14}$$

Now by an easy use of (3.8) and Lemma 0 of Mammen (1993),

$$E[c^T\boldsymbol{x}_i]^2\|\boldsymbol{x}_i\|^4 e_i^2 \leq [E(c^T\boldsymbol{x}_i)^4 e_i^2]^{1/2}[E\|\boldsymbol{x}_i\|^8 e_i^2]^{1/2} = O(p^2).$$

By using this and Lemma 4.1, at (4.14) we have $O_P((\frac{p^{1/2+\delta/4}}{n^{1/2}})^{2(h-1)}p^2 n^{-2}) = o_P(1)$. We are left with the first term in (4.13). We use the same technique as above, that is, writing one of the A as in (4.12) and thus reducing the power of A. The above scheme is used repeatedly for the rest of the calculations. For example, in the next step, for the first term in (4.13), we have

$$\begin{aligned} &n^{-3}\sum_{i=1}^{n}[c^T A_i A^{h-1}\boldsymbol{x}_i]^2 e_i^2 \\ &\leq 2n^{-3}\sum_{i=1}^{n}\left[\left(c^T\frac{A_i}{n}A_i A^{h-2}\boldsymbol{x}_i\right)^2 e_i^2 + \left(c^T\frac{A_i}{n}\boldsymbol{x}_i\boldsymbol{x}_i^T A^{h-2}\boldsymbol{x}_i\right)^2 e_i^2\right]. \end{aligned}$$

Now since A_i and $\boldsymbol{x}_i$ are independent, and $\lambda_{\text{amax}}(n^{-1}A_i) = O_P(p^{1/2+\delta/4}n^{-1/2})$, the second term is obtained to be $o_P(1)$, using similar calculations as in (4.14), based on (3.8) and Lemma 0. The term left in the above expression is $n^{-1}\sum_{i=1}^{n}(c^T(\frac{A_i}{n})^2 A^{h-2}\boldsymbol{x}_i)^2 e_i^2$. Repeating this process, ultimately we have to show $n^{-1}\sum_{i=1}^{n}(c^T(\frac{A_i}{n})^h\boldsymbol{x}_i)^2 e_i^2 = o_P(1)$, which follows easily using (3.8). This concludes the proof of (4.8).

For (4.9) and (4.10) we show the results for $h = 1$ and $h \geq 2$ separately. The result for $h = 1$ follows from direct calculation, and for higher h we employ the technique of reducing the power of A that we used in the above proof.

PROOF OF (4.9) FOR $h = 1$. We have

$$\begin{aligned} n^{-1/2}\sigma_n{}^{-1}c^T\left[(A^{*h} - A^h)\sum_{i=1}^{n}(w_i - 1)\boldsymbol{x}_i e_i\right] &= n^{-1/2}\sum_{i=1}^{n}W_i[c^T A_B\boldsymbol{x}_i]e_i \\ &= -n^{-1/2}n^{-1}\sigma_n\sum_{i=1}^{n}\sum_{j=1}^{n}W_i W_j[c^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{x}_i]e_i. \end{aligned}$$

A direct calculation shows that we have $E_B[n^{-1/2}n^{-1}\sigma_n\sum_{i=1}^{n}\sum_{j=1}^{n}W_i W_j[c^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{x}_i]e_i]^2$ is $O_P(\sigma_n^2 p^2 n^{-1})$. This is $o_P(1)$ under any one of the three alternative conditions of the theorem.

PROOF OF (4.10) FOR $h = 1$. As earlier

$$n^{-1/2}\sigma_n{}^{-1}c^T(A^{*h} - A^h)\sum_{i=1}^{n}\boldsymbol{x}_i e_i = n^{-1/2}n^{-1}\sigma_n{}^{-1}c^T(A_B)\sum_{i=1}^{n}\boldsymbol{x}_i e_i \tag{4.15}$$

$$= -n^{-1/2}n^{-1}\sum_{j=1}^{n} W_j c^T \boldsymbol{x}_j \boldsymbol{x}_j^T \left(\sum_{i=1}^{n} \boldsymbol{x}_i e_i\right).$$

Again a direct calculation shows that the term $E_B[n^{-1/2}n^{-1}\sum_{j=1}^{n} W_j c^T \boldsymbol{x}_j \boldsymbol{x}_j^T(\sum_{i=1}^{n} \boldsymbol{x}_i e_i)]^2$ is $O_P(p/n)$. Thus (4.9) and (4.10) are proved for $h = 1$, completing the proof of the Lemma when (3.12) holds since $\delta \geq 1$ implies $K \leq 2$. We now concentrate on (4.9) and (4.10) under (3.13) or (3.14) and $h > 1$.

PROOF OF (4.9) FOR $h > 1$. Since $A^* = A + A_B$, we have

$$(4.16)\qquad n^{-1/2}\sigma_n{}^{-1}c^T A^{*h}\sum_{i=1}^{n}(w_i - 1)\boldsymbol{x}_i e_i = n^{-1/2}c^T A A^{*(h-1)}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i + n^{-1/2}c^T A_B A^{*(h-1)}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i.$$

Repeating this process h times, we have

$$(4.17)\qquad n^{-1/2}\sigma_n{}^{-1}c^T A^{*h}\sum_{i=1}^{n}(w_i - 1)\boldsymbol{x}_i e_i = \sum_{j=0}^{h-1} c^T A^j A_B A^{*(h-j-1)} n^{-1/2}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i + c^T A^h n^{-1/2}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i.$$

The last term in (4.17) cancels out with the identical term left from (4.9). For the rest of the terms in (4.17), for any $j = 0, \ldots, h-1$ we have

$$\begin{aligned}
&E_B\left|n^{-1/2}c^T A^j A_B A^{*(h-j-1)}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i\right| \\
&\leq [E_B\|c^T A^j A_B\|^2]^{1/2}\left[E_B\left\|A^{*(h-1)}n^{-1/2}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i\right\|^2\right]^{1/2} \\
&\leq \lambda_{\mathrm{amax}}(A^j)[E_B\lambda_{\mathrm{amax}}(A_B^2)]^{1/2}\left[E_B\lambda_{\mathrm{amax}}(A^{*2(h-j-1)})\left\|n^{-1/2}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i\right\|^2\right]^{1/2} \\
&\leq [E_B\lambda_{\mathrm{amax}}(A_B^2)]^{1/2}\left[E_B\left\|n^{-1/2}\sum_{i=1}^{n} W_i \boldsymbol{x}_i e_i\right\|^2\right]^{1/2} \\
&= O_P(\sigma_n p^{5/4} n^{-1/2}).
\end{aligned}$$

With the second set of conditions in Theorem 3.1, that is, $p^4/n^3 \to 0$ and $\sigma_n^2 = O(n^{-1})$ the above is $o_P(1)$. With the third set of conditions also the same result holds.

PROOF OF (4.10) FOR $h > 1$. The calculations for this part is slightly more elaborate but in spirit similar to what was used for (4.9). Note that the expression in (4.10) can be expanded by using $A^* = A + A_B$, so every term in the expansion has A and A_B in various combinations. The terms without any A_B factors cancel out. The terms that involve two or more A_B matrices can be shown to be $o_P(1)$ easily. This leaves all the terms that have a single A_B matrix in them which are also $o_P(1)$. Thus

$$\begin{aligned}(4.18)\quad & n^{-1/2}\sigma_n{}^{-1}c^T A^{*h}\sum_{i=1}^n \boldsymbol{x}_i e_i \\ &= n^{-1/2}\sigma_n{}^{-1}c^T A A^{*(h-1)}\sum_{i=1}^n \boldsymbol{x}_i e_i + n^{-1/2}\sigma_n{}^{-1}c^T A_B A^{*(h-1)}\sum_{i=1}^n \boldsymbol{x}_i e_i.\end{aligned}$$

We first deal with the second term in (4.18). This is split into two terms, thus

$$\begin{aligned}(4.19)\quad & n^{-1/2}\sigma_n{}^{-1}c^T A_B A^{*(h-1)}\sum_{i=1}^n \boldsymbol{x}_i e_i \\ &= n^{-1/2}\sigma_n{}^{-1}c^T A_B A A^{*(h-2)}\sum_{i=1}^n \boldsymbol{x}_i e_i + n^{-1/2}\sigma_n{}^{-1}c^T A_B^2 A^{*(h-2)}\sum_{i=1}^n \boldsymbol{x}_i e_i.\end{aligned}$$

First take the second term in (4.19).

$$\begin{aligned} & E_B\left|n^{-1/2}\sigma_n{}^{-1}c^T A_B^2 A^{*(h-2)}\sum_{i=1}^n \boldsymbol{x}_i e_i\right| \\ &\le [\sigma_n{}^{-2}E_B\lambda_{\mathrm{amax}}(A_B^4)]^{1/2}\left[E_B\lambda_{\mathrm{amax}}(A^{*2(h-2)})\left\|n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i e_i\right\|^2\right]^{1/2} \\ &= O_P(\sigma_n{}^{-1}\sigma_n{}^2 p^2 n^{-1}) \\ &= o_P(1).\end{aligned}$$

Proceeding in this way, the terms from (4.18) and (4.19) that have two or more A_B terms in them can be shown to be small, so we are eventually left to show

$$(4.20)\qquad E_B\left[n^{-1/2}\sigma_n{}^{-1}c^T A^a A_B A^{(h-a-1)}\sum_{i=1}^n \boldsymbol{x}_i e_i\right]^2 = o_P(1)$$

for $a = 0, \ldots, h-1$. For convenience, set $\eta_1 = A^a c$ and $\eta_2 = n^{-1/2}A^{(h-a-1)}\sum_{i=1}^n \boldsymbol{x}_i e_i$. Thus (4.20) reduces to showing $\sigma_n{}^{-1}\eta_1^T A_B \eta_2 = o_P(1)$.

$$\begin{aligned} E_B[\sigma_n{}^{-1}\eta_1^T A_B\eta_2]^2 &= n^{-2}E_B\left[\sum_{i=1}^n W_i\eta_1^T \boldsymbol{x}_i\boldsymbol{x}_i^T\eta_2\right]^2 \\ &= n^{-2}(1-c_{11})\sum_{i=1}^n[\eta_1^T\boldsymbol{x}_i\boldsymbol{x}_i^T\eta_2]^2 + n^{-2}c_{11}\left[\sum_{i=1}^n \eta_1^T\boldsymbol{x}_i\boldsymbol{x}_i^T\eta_2\right]^2.\end{aligned}$$

Consider the second term.

$$c_{11}\left[n^{-1}\sum_{i=1}^n \eta_1^T\boldsymbol{x}_i\boldsymbol{x}_i^T\eta_2\right]^2 = c_{11}\left[\eta_1^T\left(n^{-1}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T - \boldsymbol{I}\right)\eta_2 + \eta_1^T\eta_2\right]^2$$

$$
\begin{aligned}
&= c_{11}[\eta_1^T A \eta_2 + \eta_1^T \eta_2]^2 \\
&\le 2c_{11}\left[\left(c^T A^h n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i e_i\right)^2 + \left(c^T A^{h-1} n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i e_i\right)^2\right] \\
&= O_P(p/n).
\end{aligned}
$$

The last step follows from Lemma 2 of Mammen (1993) where it is shown that $(c^T A^h n^{-1/2} \sum_{i=1}^n \boldsymbol{x}_i e_i)^2 = O_P(p)$, for $h = 1, \ldots, K-1$, and by using $c_{11} = O(n^{-1})$. We now show $n^{-2}\sum_{i=1}^n [\eta_1^T \boldsymbol{x}_i \boldsymbol{x}_i^T \eta_2]^2 = o_P(1)$. For this term, we have $n^{-2}\sum_{i=1}^n [\eta_1^T \boldsymbol{x}_i \boldsymbol{x}_i^T \eta_2]^2 = n^{-2}\sum_{i=1}^n [\eta_1^T \boldsymbol{x}_i]^2 [\boldsymbol{x}_i^T \eta_2]^2$. We will show the following

(i) $E[\eta_1^T \boldsymbol{x}_i]^2 = O(1)$

(ii) $E[\eta_2^T \boldsymbol{x}_i]^2 = O(p)$

and the proof will be complete. Consider (i) first. This is true if $a = 0$, so we look at $a \ge 1$ case. Then

$$
\begin{aligned}
E[c^T A^a \boldsymbol{x}_i]^2 &= n^{-2} E[c^T(\boldsymbol{x}_i \boldsymbol{x}_i^T + A_i) A^{a-1} \boldsymbol{x}_i]^2 \\
&\le 2n^{-2} E[c^T \boldsymbol{x}_i \boldsymbol{x}_i^T A^{a-1} \boldsymbol{x}_i]^2 + 2n^{-2} E[c^T A_i A^{a-1} \boldsymbol{x}_i]^2.
\end{aligned}
$$

For the first term,

$$
\begin{aligned}
n^{-2} E[c^T \boldsymbol{x}_i \boldsymbol{x}_i^T A^{a-1} \boldsymbol{x}_i]^2 &= n^{-2}[E(c^T \boldsymbol{x}_i)^4]^{1/2}[E(\boldsymbol{x}_i^T A^{a-1} \boldsymbol{x}_i)^4]^{1/2} \\
&= O_P(p^2 n^{-2}) \lambda_{\mathrm{amax}}(A^{2(a-1)}) \\
&= o_P(1).
\end{aligned}
$$

The term $E[c^T \frac{A_i}{n} A^{a-1} \boldsymbol{x}_i]^2$ is treated by breaking up A again if $a - 1 \ge 1$, and using the independence of A_i and $\boldsymbol{x}_i$. This proves (i). For (ii), write $b = h - a - 1$ and since the relation easily holds for $b = 0$, we look at $b \ge 1$. Note that

$$
\left[\boldsymbol{x}_i^T A^b n^{-1/2}\sum_{j=1}^n \boldsymbol{x}_j e_j\right]^2 \le 2[\boldsymbol{x}_i^T A^b n^{-1/2} \boldsymbol{x}_i e_i]^2 + 2\left[\boldsymbol{x}_i^T A^b n^{-1/2}\sum_{j \ne i} \boldsymbol{x}_j e_j\right]^2.
$$

The first term easily seen to be $O_P(p^2 n^{-1}) = o_P(p)$. We use the notation $\eta_i = n^{-1/2}\sum_{j\ne i} \boldsymbol{x}_j e_j$ for the remaining part of the proof. Then the second term of the above expression is $E[\boldsymbol{x}_i^T A^b \eta_i]^2$. Using independence of $\boldsymbol{x}_i$ and η_i this is seen to be $O(p)$ if $b = 0$, so take $b \ge 1$. Then

$$
E[\boldsymbol{x}_i^T A^b \eta_i]^2 \le 2n^{-2} E[\boldsymbol{x}_i^T A^{b-1} \boldsymbol{x}_i]^2 [\boldsymbol{x}_i^T \eta_i]^2 + 2E\left[\boldsymbol{x}_i^T A^{b-1} \frac{A_i}{n} \eta_i\right]^2.
$$

Using a Cauchy-Schwartz inequality and independence of $\boldsymbol{x}_i$ and η_i the first term is seen to be $o(p)$ and the second term is treated by techniques similar to ones used earlier. □

PROOF OF LEMMA 4.2. Note that the bootstrap and the original estimators are given by $\hat{\beta}_B = (\sum_{i=1}^n \boldsymbol{x}_i^* \boldsymbol{x}_i^{*T})^{-1} \sum_{i=1}^n \boldsymbol{x}_i^* y_i^*$ and $\hat{\beta} = (\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1} \sum_{i=1}^n \boldsymbol{x}_i y_i$. Since $E_B \lambda_{\mathrm{amax}}(A^*) = o_P(1)$, we can write $(\sum_{i=1}^n \boldsymbol{x}_i^* \boldsymbol{x}_i^{*T})^{-1} = n^{-1}(\boldsymbol{I} - A^*)^{-1} = n^{-1}[\boldsymbol{I} + A^* + \cdots]$

and hence $\hat{\beta}_B = \beta + n^{-1}\sum_{h\geq 0} A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^*$ and $\hat{\beta} = \beta + n^{-1}\sum_{h\geq 0} A^h \sum_{i=1}^n \boldsymbol{x}_i e_i$ Thus

$$\begin{aligned} n^{1/2}{\sigma_n}^{-1}(\hat{\beta}_B - \hat{\beta}) &= n^{-1/2}\sum_{h\geq 0} A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* - n^{-1/2}\sum_{h\geq 0} A^h \sum_{i=1}^n \boldsymbol{x}_i e_i \\ &= n^{-1/2}\sum_{i=1}^n W_i c^T \boldsymbol{x}_i e_i + n^{-1/2}\sum_{h=1}^{K-1} A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \\ &\quad - n^{-1/2}\sum_{h=1}^{K-1} A^h \sum_{i=1}^n \boldsymbol{x}_i e_i + n^{-1/2}\sum_{h\geq K} A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \\ &\quad - n^{-1/2}\sum_{h\geq K} A^h \sum_{i=1}^n \boldsymbol{x}_i e_i. \end{aligned}$$

From Lemma 4.4, it follows that

$$n^{-1/2}\sum_{h=1}^{K-1} c^T A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* - n^{-1/2}\sum_{h=1}^{K-1} c^T A^h \sum_{i=1}^n \boldsymbol{x}_i e_i = r_{1nB}$$

where $P_B[|r_{1nB}| > \epsilon] = o_P(1)$ for every $\epsilon > 0$. Note that in Lemma 3 of Mammen (1993), it has already been shown $n^{-1/2}\sum_{h\geq K} c^T A^h \sum_{i=1}^n \boldsymbol{x}_i e_i = o_P(1)$. Now let $n^{-1/2}\sum_{h\geq K} c^T A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* = r_{2nB}$, then we now show $P_B[|r_{2nB}| > \epsilon] = o_P(1)$ to complete the proof. We use the technique of Lemma 3 of Mammen (1993), thus for almost all data sequences

$$\left| n^{-1/2}\sum_{h\geq K} c^T A^{*h}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \right| \leq \lambda_{\text{amax}}(A^{*K})\frac{1}{1-\lambda_{\text{amax}}(A^*)}\left\| n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \right\|.$$

So

$$P_B[|r_{2nB}| > \epsilon] \leq \epsilon^{-2} E_B r_{2nB}^2$$

$$\leq \epsilon^{-2}[E_B \lambda_{\text{amax}}(A^{*2K})]\left[E_B \left\| n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \right\|\right] \tag{4.21}$$

$$\leq \epsilon^{-2}[E_B \lambda_{\text{amax}}(A^{*4})]\left[E_B \left\| n^{-1/2}\sum_{i=1}^n \boldsymbol{x}_i^* e_i^* \right\|\right] \quad \text{if} \quad \delta < 2. \tag{4.22}$$

At (4.21) we can separate the cases where $\delta \geq 2$ and $\delta < 2$. Notice that when we have $E_B \lambda_{\text{amax}}(A^*) = O_P(p^{1/2+\delta/4} n^{-1/2})$, then (4.21) is $o_P(1)$, the calculations being identical with that of Lemma 3 of Mammen (1993). If $\delta \geq 2$, then $K = 1$ and from Lemma 4.1, we have $E_B \lambda_{\text{amax}}(A^*) = O_P(p^{1/2+\delta/4} n^{-1/2})$. Hence we need to look at $\delta < 2$ case under the condition $E_B \lambda_{\text{amax}}(A^{*4}) = O_P(\sigma_n^4 p^3 n^{-2})$. For this case, we use (4.22) and get $P_B[|r_{2nB}| > \epsilon] = O_P({\sigma_n}^4 n^{-2} p^4)$. The last relation is $o_P(1)$ under any one of the three sets of possible conditions on σ_n^2 and δ given in the theorem. □

Acknowledgements

The authors are grateful to the referees for their insightful comments. This has lead to a clearer presentation of our results. We are also grateful to them for pointing out a few crucial references.

References

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

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

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

Chatterjee, S. and Bose, A. (2000*a*). Variance estimation in high dimensional regression models, *Statist. Sinica*, **10**, 497–515.

Chatterjee, S. and Bose, A. (2000*b*). Generalised bootstrap for solutions of martingale estimating equations (preprint).

Haberman, S. J. (1977*a*). Log-linear and frequency tables with small expected cell counts, *Ann. Statist.*, **5**, 1148–1169.

Haberman, S. J. (1977*b*). Maximum likelihood estimates in exponential response models, *Ann. Statist.*, **5**, 815–841.

Hall, P. and Mammen, E. (1994). On general resampling algorithms and their performance in distribution estimation, *Ann. Statist.*, **22**, 2011–2030.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

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

Mammen, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap, *Ann. Statist.*, **17**, 382–400.

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

Mammen, E. (1996). Empirical process of residuals for high dimensional linear models, *Ann. Statist.*, **24**, 307–335.

Mason, D. A. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap, *Ann. Statist.*, **20**, 1611–1624.

Portnoy, S. (1984). Asymptotic behaviour of M-estimators of p regression parameters when p^2/n is large, I: Consistency, *Ann. Statist.*, **12**, 1298–1309.

Portnoy, S. (1985). Asymptotic behaviour of M-estimators of p regression parameters when p^2/n is large, II: Normal approximation, *Ann. Statist.*, **13**, 1403–1417.

Portnoy, S. (1988). Asymptotic behaviour of the empirical distribution of M-estimator residuals from a regression model with many parameters, *Ann. Statist.*, **14**, 1152–1170.

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

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

Sauermann, W. (1989). Bootstrapping the maximum likelihood estimator in high-dimensional log-linear models, *Ann. Statist.*, **17**, 1198–1216.

Shorack, G. (1982). Bootstrapping robust regression, *Comm. Statist. Theory Methods*, **11**, 961–972.

Wu, C. F. J. (1990). On the asymptotic properties of the jackknife histogram, *Ann. Statist.*, **18**, 1438–1452.

Yin, Y. Q., Bai, D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix, *Probab. Theory and Related Fields*, **78**, 509–521.

Zheng, Z. G. and Tu, D. S. (1988). Random weighting methods in regression models, *Sci. Sinica Ser. A*, **31**, 1442–1459.